



HAL
open science

Régulation de la formule génomique d'un nanovirus dans sa plante hôte

Mélia Bonnamy

► **To cite this version:**

Mélia Bonnamy. Régulation de la formule génomique d'un nanovirus dans sa plante hôte. Micro et nanotechnologies/Microélectronique. Université de Montpellier, 2023. Français. NNT : 2023UMONG013 . tel-04622522

HAL Id: tel-04622522

<https://theses.hal.science/tel-04622522>

Submitted on 24 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En biologie des interactions – mécanismes des interactions parasitaires pathogènes et symbiotiques

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

Unité de recherche MIVEGEC – Maladies Infectieuses et Vecteurs : Ecologie, Génétique, Evolution et Contrôle
Unité de recherche PHIM – Plant Health Institute Montpellier

Régulation de la formule génomique d'un nanovirus dans sa plante hôte

Présentée par Mélia BONNAMY

Le 14 septembre 2023

Sous la direction de Yannis MICHALAKIS
Et le co-encadrement de Stéphane BLANC

Devant le jury composé de

Emanuela NORIS, Senior Researcher, Italian National Research Council

Hervé VANDERSCHUREN, Professeur, Katholieke Universiteit Leuven

Ioana POPESCU, Maître de conférences, Université Paris-Saclay

Anne-Sophie GOSSELIN-GRENET, Professeure des universités, Université de Montpellier

David GILMER, Professeur, Université de Strasbourg

Rapporteure

Rapporteur

Examinatrice

Examinatrice

Examineur



UNIVERSITÉ
DE MONTPELLIER

Les virus multipartites se caractérisent par leur génome divisé en plusieurs segments génomiques, chacun encapsidé individuellement dans une particule virale. Une particularité permise par la nature segmentée de leur génome est l'accumulation inégale des différents segments génomiques dans les tissus infectés. Le pattern reproductible et spécifique d'accumulation des segments génomiques au sein de l'hôte est désigné « formule génomique ». Quelques résultats expérimentaux laissent supposer que la formule génomique joue un rôle fonctionnel de régulation de l'expression des gènes viraux en modulant leur nombre de copies. Par ailleurs, les phénomènes qui la régulent étaient jusqu'alors entièrement inconnus. Ainsi, en prenant pour modèle le nanovirus octopartite faba bean necrotic stunt virus (FBNSV ; famille *Nanoviridae*), cette thèse avait pour objectif d'identifier les mécanismes qui façonnent la formule génomique.

Nous avons d'abord recensé les connaissances sur les mécanismes de réplication des *Nanoviridae* disponibles dans la littérature, afin d'y déceler des informations indiquant une réplication différentielle des segments génomiques qui pourraient expliquer la formule génomique. Les données sur la réplication des nanovirus sont peu nombreuses et très souvent déduites des connaissances sur les *Geminiviridae* qui sont, comme les *Nanoviridae*, des phytovirus à ADN simple-brin. Nous avons mis en évidence de nombreux points d'ombre dans les connaissances sur les mécanismes de réplication des *Nanoviridae* et *Geminiviridae*. Nos recherches n'ont pas donné de réponse formelle à notre questionnement, mais ont permis d'émettre des hypothèses quant à l'implication des différents modes de réplication de ces virus dans la réplication et l'expression génique virale.

Dans une seconde partie, nous avons expérimentalement démontré que les deux segments génomiques du FBNSV les plus fréquents dans *V. faba* possèdent une plus grande capacité intrinsèque à s'accumuler dans les tissus de cet hôte que les autres segments. Néanmoins, cette régulation à l'échelle du segment n'est pas suffisante pour générer la formule génomique, suggérant l'implication de mécanismes supplémentaires agissant au niveau supra-segmentaire. De manière cohérente, l'absence de certains segments génomiques au cours d'une infection systémique impacte l'accumulation des autres segments, indiquant que la formule génomique est une propriété du groupe de segments. L'ensemble de ces résultats démontre que la formule génomique du FBNSV est façonnée par un processus complexe agissant à la fois au niveau du segment individuel et du groupe de segments.

Enfin, il est établi que la formule génomique du FBNSV change drastiquement d'un hôte à un autre, indiquant que celle-ci est déterminée par le génotype de la plante. Par ailleurs, l'importance de l'état physiologique de l'hôte sur la formule génomique n'a jusqu'à présent pas été documenté. Après avoir déterminé la formule génomique du FBNSV dans des plantes cultivées dans différentes conditions abiotiques et biotiques, nos premiers résultats indiquent que, de manière surprenante, la formule génomique semble peu impactée par l'état physiologique de la plante.

En conclusion, cette thèse a permis d'établir que les mécanismes de réplication des virus à ADN simple-brin de plantes pourraient potentiellement être impliqués dans la régulation du nombre de copies des gènes viraux. La formule génomique du FBNSV résulte de processus agissant à la fois au niveau du segment individuel et au niveau supra-segment. Enfin, la formule génomique est sujette à de forts changements lors de passages d'une espèce hôte à une autre, mais elle semble globalement robuste aux modifications de l'état physiologique de l'hôte.

Mots-clés : virus multipartite, nanovirus, formule génomique, *copy number variation*, *rolling-circle replication*, *recombination-dependent replication*

Multipartite viruses are characterized by their genome divided into several genome segments, each individually encapsidated in a viral particle. A particular feature of the segmented nature of their genomes is the unequal accumulation of different genome segments in infected tissues. The reproducible and host specific pattern of accumulation of genome segments is referred to as the "genome formula". Some experimental results suggest that the genome formula plays a functional role in regulating viral gene expression by modulating gene copy number. However, the mechanisms regulating the genomic formula were entirely unknown. Using the octopartite faba bean necrotic stunt virus (FBNSV; family *Nanoviridae*) as a model, the objective of this thesis was to identify the mechanisms that shape the genome formula.

We began by reviewing the literature on the replication mechanisms of *Nanoviridae*, looking for evidence of differential replication of genome segments that might explain the genome formula. Data on nanovirus replication is scarce and very often deduced by analogy from results on the *Geminiviridae* which, like the *Nanoviridae*, are single-stranded DNA phytoviruses. We have highlighted many gaps in our understanding of the replication mechanisms of *Nanoviridae* and *Geminiviridae*. Our literature review did not provide a formal explanation for our initial questioning, but enabled us to propose hypotheses on the involvement of the different replication modes of these viruses in viral replication and gene expression.

In a second part, we have experimentally demonstrated that the two most frequent FBNSV genome segments in *Vicia faba* have a greater intrinsic capacity to accumulate in the tissues of this host than the other segments. However, this segment-level regulation is not sufficient to generate the genome formula, suggesting the involvement of additional mechanisms acting at the supra-segment level. Consistently, the absence of certain genome segments during systemic infection affects the accumulation of other segments, indicating that the genome formula is a property of the segment group. Taken together, these results demonstrate that the FBNSV genome formula is shaped by a complex process acting at both the level of the individual segment and the segment group.

Finally, it has been established that the FBNSV genome formula changes drastically from one host to another, indicating that it is determined by the plant genotype. Furthermore, the importance of the host's physiological state on the genome formula has so far not been investigated. Having determined the genome formula of FBNSV in plants grown under different abiotic and biotic conditions, our initial results indicate that, surprisingly, the genome formula appears to be little affected by the physiological state of the plant.

In conclusion, this thesis has established that the replication mechanisms of plant single-stranded DNA viruses could potentially be involved in gene copy number variation. The genome formula of the FBNSV results from processes acting at both the individual segment and supra-segment levels. Finally, the genome formula is subject to strong changes when moving from one host species to another, but overall appears to be robust to changes in the host's physiological state.

Key words : multipartite viruses, nanovirus, genome formula, copy number variation, rolling-circle replication, recombination-dependent replication

Table des matières

Introduction	5
I. Les virus multipartites.....	5
A. Constituants d'un virus	5
B. Architecture et <i>packaging</i> des génomes viraux.....	7
C. Coûts du multipartitisme	9
1. Un nombre de particules virales transmises et initiant une nouvelle infection suffisamment grand	9
2. Mouvements groupés	11
3. Complémentation inter-cellulaire	13
4. Acquisition séparée des différents segments génomiques.....	13
D. Avantages potentiels du multipartitisme	15
1. Réplication.....	15
2. Stabilité des particules	15
3. Expression des gènes viraux.....	17
4. Réassortiments génétiques.....	19
E. Conclusion sur les coûts et avantages du multipartitisme.....	19
II. La formule génomique	21
A. Définition.....	21
B. Rôle fonctionnel de la formule génomique	23
C. Etablissement de la formule génomique	25
III. Le FBNSV pour étudier la formule génomique	29
A. Généralités sur les <i>Nanoviridae</i>	29
1. Description	29
2. Gamme d'hôtes.....	29
3. Répartition géographique et importance économique	31
4. Les alpha-satellites.....	31
B. Notre système d'étude FBNSV – <i>Vicia faba</i>	33
C. Les protéines codées par le FBNSV	35
D. Transmission du FBNSV	39
IV. Problématique et objectifs de thèse	41

Chapitre 1 : Etude bibliographique : la réplication des nanovirus et sa potentielle implication dans l'établissement de la formule génomique	43
I. Contexte	43
II. Article de revue : Replication mechanisms of circular ssDNA plant viruses and their potential implication in viral gene expression regulation	44
Chapitre 2 : Etude des mécanismes régulant la formule génomique	83
I. Contexte	83
II. Article : The genome formula of a multipartite virus is regulated both at the individual segment and the segment group levels.....	84
III. Influence des interactions entre les différents composants viraux dans la mise en place de la formule génomique	119
A. Objectif.....	119
B. Méthode.....	120
1. Optimisation d'un protocole de préparation, transfection et culture de protoplastes de fèves	120
2. Transfection du génome viral	120
C. Perspectives	121
IV. Adaptabilité de la formule	122
A. Expression de protéines du FBNSV <i>in planta</i>	122
1. Choix du virus vecteur d'expression.....	123
2. Co-infections avec le FBNSV	125
3. Stratégie de clonage de séquences codantes du FBNSV dans le clone infectieux du PSbMV ..	128
4. Discussion et perspectives	128
B. Formule et <i>silencing</i> antiviral	130
1. Principe	130
2. Méthode.....	130
3. Résultats et discussion	132
Chapitre 3 : Etude de l'impact de l'état physiologique de la plante hôte sur la formule génomique	136
I. Contexte	136
II. Article : Impact of the physiological state of the host plant on the genome formula of a nanovirus	137
Conclusion générale.....	156
Références hors articles	161
Annexe 1	170

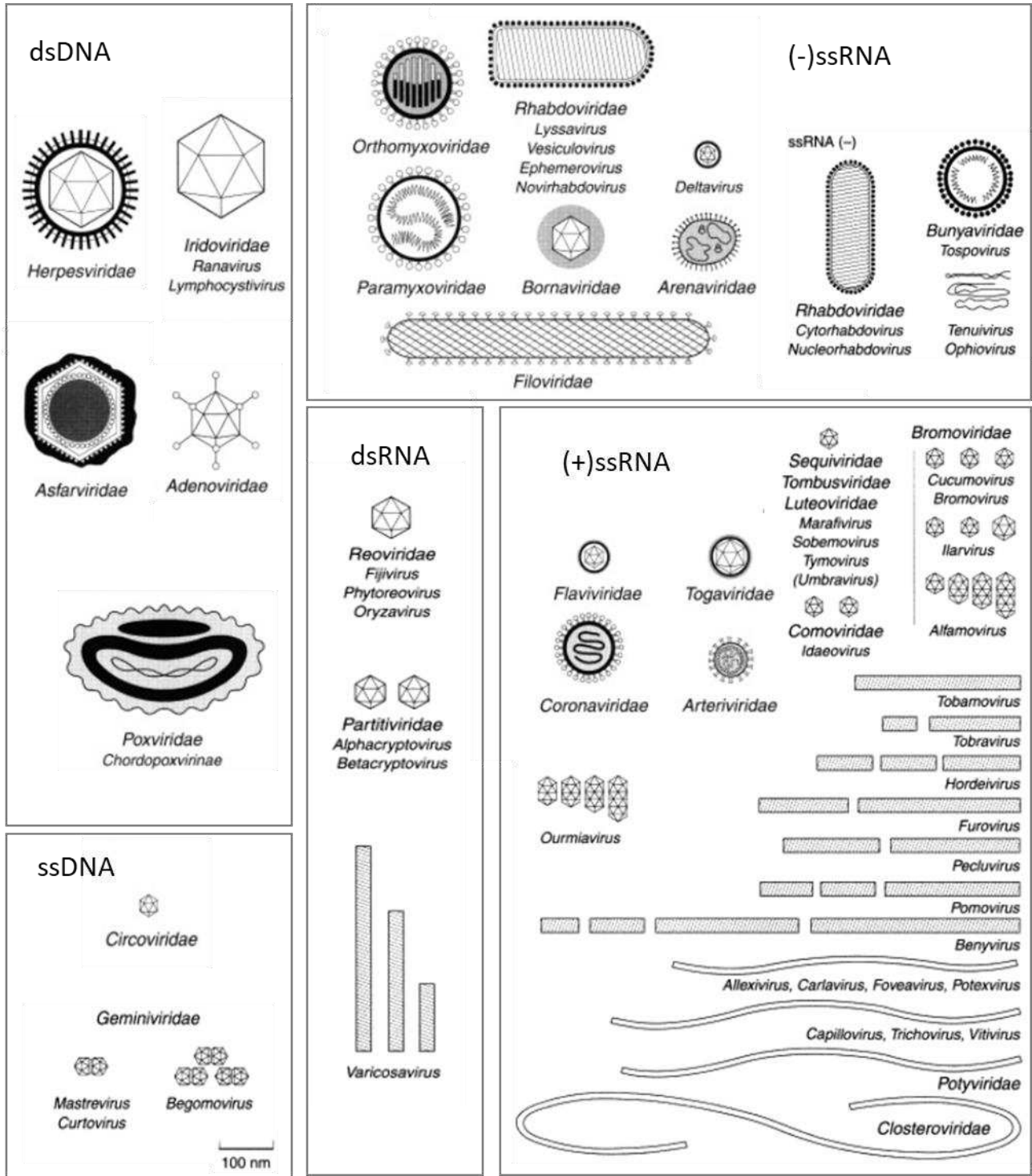


Figure 1 : Exemples de la diversité des virus. Représentation de la structure de certains virus de vertébrés et plantes, en fonction la nature de leur génome.

Adapté de Fauquet, 1999

Introduction

Les virus sont omniprésents dans notre environnement et représentent probablement les entités biologiques les plus fréquentes sur Terre. Ils sont présents dans tous les écosystèmes terrestres et peuvent infecter tous les règnes du vivant : bactéries, archées et autres organismes unicellulaires, plantes, champignons, invertébrés et vertébrés. Etant des parasites intracellulaires obligatoires, ils sont dépendants d'une cellule hôte pour effectuer leur cycle d'infection. Les infections virales peuvent parfois avoir un effet négatif sur l'hôte. Les virus sont la cause de nombreuses maladies humaines, responsables de grandes épidémies qui ont contribué, et contribuent toujours, à façonner l'histoire de nos civilisations. Les infections virales affectant des cultures agricoles ou des élevages peuvent également conduire à des pertes de production considérables et ainsi entraîner de lourdes conséquences économiques et sociales. Cependant, la majorité des virus n'ont pas d'impact sur leur hôte et d'autres sont même bénéfiques. Par exemple, certaines infections virales confèrent une résistance à l'infection par d'autres pathogènes (Skalka et al., 2015). Ainsi, les virus ne sont pas uniquement nos ennemis, mais des constituants essentiels de la vie sur Terre.

I. Les virus multipartites

A. Constituants d'un virus

Les virus sont des structures particulières composées d'un génome protégé par une capsid. Le génome viral peut être constitué d'acide désoxyribonucléique (ADN) ou d'acide ribonucléique (ARN). Le génome peut être circulaire ou linéaire, simple (ssDNA ou ssRNA) ou double brin (dsDNA ou dsRNA). Les génomes à ARN simple brin peuvent être de polarité positive ((+)ssRNA) ou négative ((-)ssRNA) : les virus (+)ssRNA peuvent être directement traduits pour synthétiser les protéines virales, alors que les (-)ssRNA doivent passer par une étape de transcription du brin positif complémentaire pour l'expression de leurs gènes (Baltimore, 1971). La capsid est composée d'un assemblage de protéines de capsides virales et peut être de symétrie icosaédrique, hélicoïdale ou mixte. Certains virus peuvent également être enveloppés. Dans ce cas, l'enveloppe est formée par des protéines d'enveloppe codées par le virus associées à des composants membranaires dérivés de la cellule hôte. L'assemblage du génome, de la

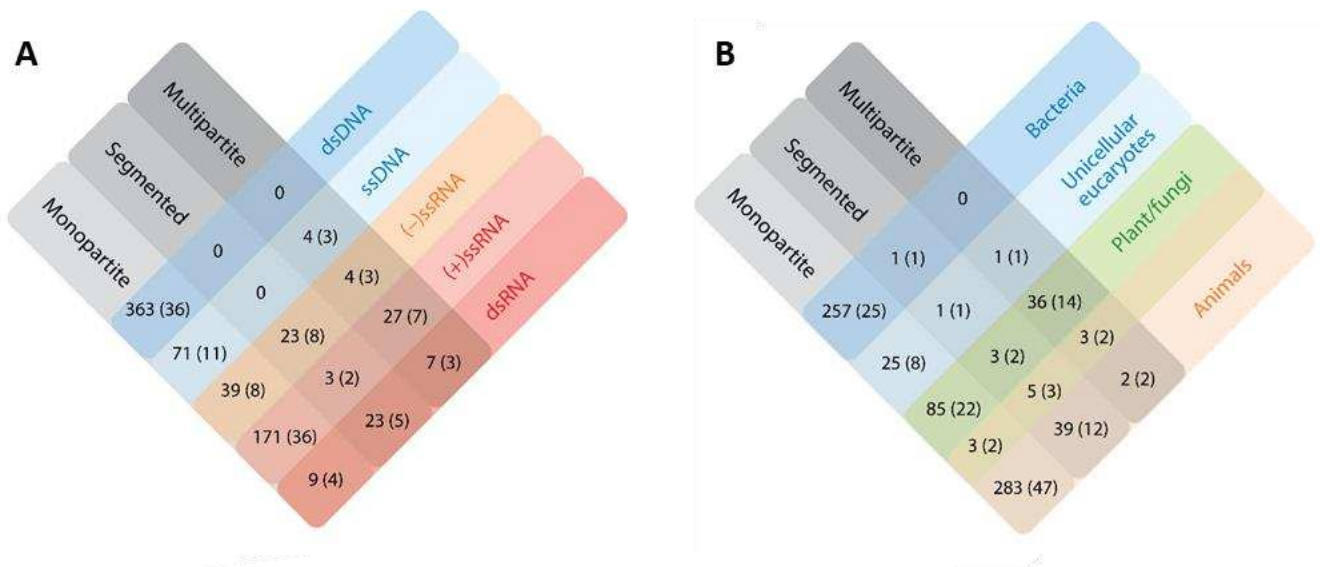


Figure 2 : Distribution des différentes architectures génomiques dans les différents types d'acides nucléiques (A) ou entre les différents types d'hôtes (B). Les nombres simples représentent le nombre de genres et ceux entre parenthèses représentent le nombre de familles. Le chevauchement des plages correspondant aux virus d'animaux et de plantes/champignons représente des virus de plantes se répliquant dans leurs vecteurs animaux. Lorsque les hôtes ne sont pas identifiés, les genres et familles correspondants ne sont pas comptés.

Adapté de Michalakis and Blanc, 2020

capside et éventuellement de l'enveloppe constitue la particule virale ou virion. Il existe une grande diversité de morphologies des particules virales, définies par la symétrie de la capsidie et la présence d'une enveloppe (**Figure 1**).

B. Architecture et *packaging* des génomes viraux

Les virus peuvent être classés en trois grandes catégories selon leur architecture génomique : monopartites, segmentés ou multipartites. Chez les virus monopartites l'information génétique est portée par une unique molécule d'ADN ou d'ARN encapsidée dans une particule virale. Chez les virus segmentés, l'information génétique est divisée en au moins deux molécules, appelées segments génomiques, regroupées dans une même particule virale. Enfin, les virus multipartites ont également leur information génétique divisée en plusieurs segments génomiques, mais chaque segment est encapsidé individuellement dans une particule virale. Ainsi, plusieurs particules sont nécessaires pour reconstituer l'intégralité du génome d'un virus multipartite.

Les virus monopartites sont majoritaires dans presque toutes les classes d'acides (ribo)nucléiques, à l'exception des virus à dsRNA où ils ne sont qu'une minorité. Les virus segmentés sont uniquement des virus à ARN. Ils représentent la majorité des virus à dsRNA. Très peu sont des virus à (+)ssRNA. Enfin, les virus multipartites sont majoritairement des virus à (+)ssRNA et peuvent également être ssDNA, (-)ssRNA ou dsRNA (**Figure 2A**) (Michalakis and Blanc, 2020).

Les différents types, architectures et modes d'encapsidation des génomes viraux ne sont pas répartis uniformément dans les différents règnes. Leur distribution a été résumée par Michalakis and Blanc, 2020. Les virus monopartites constituent la majorité des virus identifiés à ce jour. Ils sont majoritairement présents chez les bactéries et les animaux. Les virus segmentés sont majoritairement présents chez les animaux. Les virus multipartites sont très communs chez les plantes et les champignons. Aucun virus multipartite infectant des bactéries n'a été détecté à ce jour et peu l'ont été chez des animaux (**Figure 2B**).

C. Coûts du multipartitisme

La séparation de plusieurs parties du génome des virus multipartites dans plusieurs particules virales pose un problème de maintien de l'intégrité génomique lors de transmissions intra- et inter-hôte, simplement lié à l'absence de lien physique entre les ou certains gènes. Le risque de perdre une partie de l'information génétique est d'autant plus important que le nombre de segments génomiques est grand (Michalakis and Blanc, 2020). Néanmoins, en s'appuyant sur des données d'encapsulation et de transmission de différents virus à ARN au génome segmenté ou non, Dall'Ara et al., 2016 ont remarqué que les virus au génome segmenté ne semblent pas avoir plus de chance de perdre une partie essentielle du génome que les virus monopartites. Certaines conditions peuvent réduire le risque de perdre une partie de l'information génétique pour un virus multipartite.

1. Un nombre de particules virales transmises et initiant une nouvelle infection suffisamment grand

Le risque de perdre de l'information génétique peut être minimisé si le nombre de particules virales transmises est suffisamment grand pour assurer qu'au moins une copie de chaque segment soit transmise (Michalakis and Blanc, 2020). Lors de leur transmission, les populations virales subissent des goulots d'étranglement. Un goulot d'étranglement correspond à une réduction transitoire du nombre de particules virales, et peut intervenir à différentes étapes du cycle viral : au niveau intra-hôte lors de l'infection cellulaire, puis lors de la propagation systémique, et également lors de la transmission inter-hôte (transmission horizontale ou verticale) (Zwart and Elena, 2015). Les goulots d'étranglement entraînent une réduction de la diversité génétique, puisqu'ils consistent en un échantillonnage aléatoire des génomes dans la population virale source (Zwart and Elena, 2015). Ils déterminent en général la MOI (*Multiplicity of infection*) qui correspond au nombre de génomes viraux qui initient effectivement un nouveau foyer d'infection (Sicard, 2014). Chez les virus multipartites, la notion de MOI se complexifie. En effet, si le nombre de particules transmises est relativement faible, alors des génomes incomplets peuvent être transmis, des segments peuvent être manquants alors que d'autres peuvent être transmis en plusieurs copies.

Alors qu'un nombre conséquent d'études semblent converger vers une estimation d'une MOI faible pour les virus monopartites ou segmentés (Gutiérrez et al., 2012), peu de données sont disponibles concernant ces paramètres de transmission pour des virus multipartites. Betancourt et al., 2008 ont estimé que seulement 1 à 2 copies de chacun des trois segments génomiques du cucumber mosaic virus

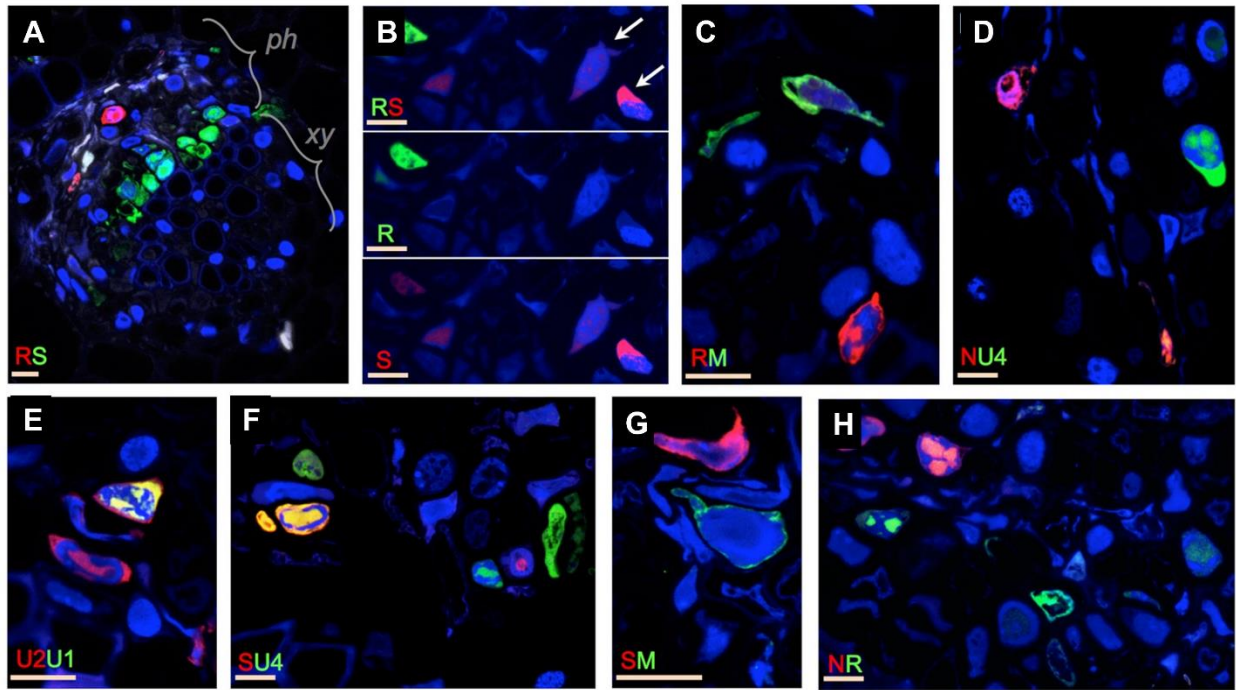


Figure 3 : Localisation cellulaire de certains segments du FBNSV dans la plante hôte de *V. faba*. Des paires de segments sont sondés par FISH vert et rouge sur des coupes transversales de pétioles de fèves infectées. Les faisceaux de phloème (ph, le seul tissu infecté par les nanovirus) et de xylème (xy) sont indiqués en (A). Les canaux de couleur sont fusionnés dans toutes les images sauf en B où le vert et le rouge sont fusionnés et séparés pour mettre en évidence l'accumulation du segment S dans les cellules où le segment R n'est pas détecté (illustré par des flèches blanches). Les noyaux sont colorés au bleu DAPI. Barres horizontales = 10 microns.

Sicard et al., 2019

(CMV ; famille *Bromoviridae* ; (+)ssRNA) initient une nouvelle infection après transmission par le puceron *Aphis gossypii* (mode de transmission non persistant). Gallet et al., 2018 ont évalué que le nombre de particules transmises pour deux segments du virus à ssDNA octopartite faba bean necrotic stunt virus (FBNSV ; famille *Nanoviridae*) pendant la transmission par le puceron vecteur *Acyrtosiphon pisum* (mode de transmission circulant non multipliant) est inférieur à 15. Enfin, Miyashita and Kishino, 2010 ont déterminé que la MOI lors de la transmission de cellule à cellule du virus à (+)ssRNA bipartite soil-borne wheat mosaic virus (SBWMV ; famille *Furovirus*) est d'environ 5 copies de chaque segment. Toutes ces études montrent que les virus multipartites subissent des goulots d'étranglement sévères lors des transmissions inter- et intra-hôte, suggérant que les échecs de transmission de cellule à cellule ou d'hôte à hôte liés à la perte d'un segment génomique doivent être très fréquents (Gallet et al., 2018).

Une étude théorique de Iranzo and Manrubia, 2012 détermine le nombre minimal de particules virales initiant une infection, soit la MOI minimale, pour un virus multipartite. Dans leur modèle, les auteurs font l'hypothèse d'un avantage sélectif relativement grand du multipartitisme et de l'équifréquence des segments génomiques. Ils ont ainsi estimé que la MOI théorique nécessaire à l'établissement d'une population multipartite, en compétition avec un génome ancestral non segmenté, augmente avec le nombre de segments : deux particules pour deux segments, environ 30 particules pour trois segments et plus de 100 particules pour quatre segments ou plus. La MOI diminue si les génomes courts sont fortement favorisés, mais reste toutefois très élevée à partir de cinq segments (100 particules). Les auteurs soulignent qu'une telle MOI paraît irréaliste pour des conditions biologiques réelles.

En conclusion, ces études concluent sur une énigme non résolue, en indiquant que le nombre des particules qui sont transmises et qui initient de nouveaux foyers d'infection est faible. Le risque de perdre une partie du génome viral lors de la transmission d'un virus multipartite ne semble donc pas être compensé par la transmission d'un grand nombre de particules.

2. Mouvements groupés

Une solution alternative à la transmission d'un grand nombre de particules virales serait que les segments voyagent ensemble (Gilmer et al., 2018), sous forme de complexes contenant chacun au moins une copie de chaque. Dall'Ara et al., 2016 font notamment l'hypothèse d'un transport intra-plante groupé des segments de phytovirus à ARN, sous forme non-encapsidée, *via* des interactions entre les différents ARNs génomiques. Bien que la transmission de certains virus sous forme de structures renfermant plusieurs virions ait été observée (Sanjuán and Thoulouze, 2019), à ce jour, aucune étude ne démontre

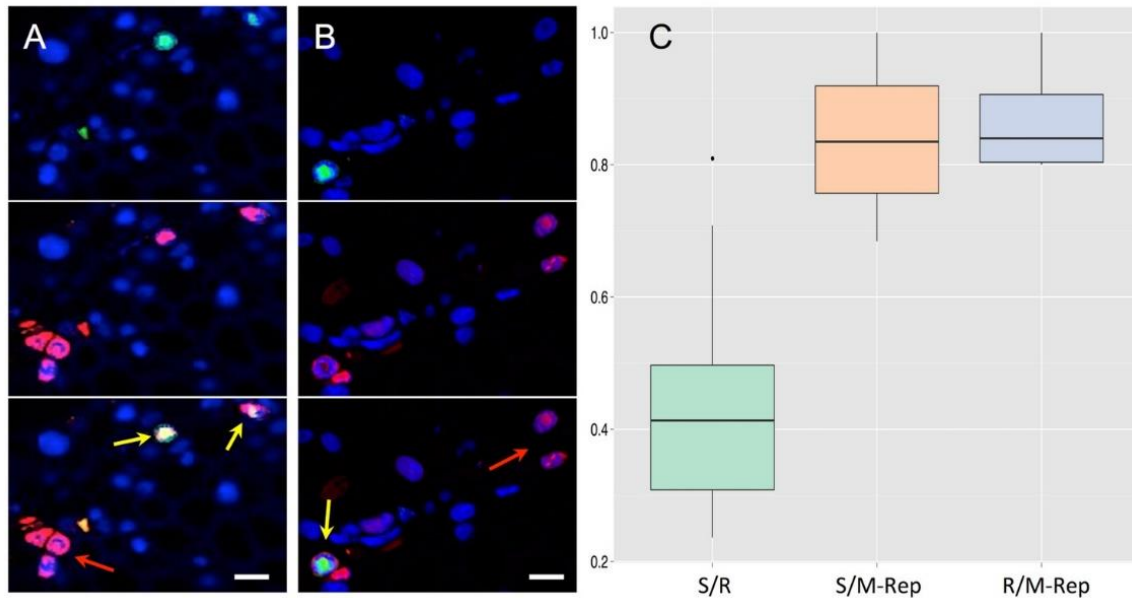


Figure 4 : Accumulation de la protéine M-Rep dans les cellules contenant le segment S ou le segment R. Le segment R (A) ou S (B) est marqué par FISH avec une sonde fluorescente verte tandis que la protéine M-Rep est immuno-marquée avec un fluorochrome rouge. Les canaux verts et rouges sont séparés (en haut et au milieu) et fusionnés (en bas). Les noyaux sont colorés au DAPI en bleu. La barre horizontale est de 10 microns. Les flèches jaunes indiquent les cellules où le segment d'ADN et la protéine M-Rep sont détectés, tandis que les flèches rouges illustrent les cas où seule la protéine M-Rep est détectable. Dans le diagramme en boîte à moustache ou boxplot (C), la proportion de cellules avec le segment S contenant également le segment R (graphique vert à gauche, estimé à partir de 10 pétioles) est comparée à la proportion de cellules avec le segment S contenant également la protéine M-Rep (graphique orange au milieu, estimé à partir de 14 pétioles). La proportion de cellules individuelles avec un segment R détectable contenant également la protéine M-Rep a été estimée à partir de 4 pétioles infectés (graphique bleu à droite). La différence entre le diagramme vert de gauche et le diagramme orange du milieu est très significative.

Sicard et al., 2019

expérimentalement l'existence de structures permettant le mouvement groupé (inter- ou intra-hôte) des différents segments génomiques de virus multipartites.

3. Complémentation inter-cellulaire

En localisant des segments génomiques du FBNSV dans des tissus de fèves infectées, Sicard et al., 2019 ont mis en évidence que ceux-ci colocalisent rarement dans une même cellule (**Figure 3**). Les segments s'accumulent de manière indépendante dans des cellules distinctes, y compris en absence du segment génomique R codant pour la protéine Rep nécessaire à la réplication (**Figure 3**). De plus, la protéine Rep est souvent présente dans des cellules où le segment R n'est pas détecté, suggérant qu'au moins certains des produits d'expression des segments (ARNm et/ou protéines) peuvent se déplacer de cellule à cellule (**Figure 4**). La colocalisation dans certaines cellules du segment S et de la protéine Rep, mais pas du segment R, confirme ce phénomène et permet une complémentation du système viral entre cellules voisines (**Figure 4**). Ce fonctionnement multicellulaire du FBNSV réduit le coût de la segmentation du génome à l'échelle intra-hôte (Sicard et al., 2019).

4. Acquisition séparée des différents segments génomiques

Di Mattia et al., 2022 mettent en évidence que le mouvement concomitant des huit segments du FBNSV d'une plante infectée vers une plante receveuse n'est pas nécessaire à la réussite de l'infection. D'une part, un puceron vecteur est capable d'acquérir une partie du génome du virus sur une première plante, puis d'acquérir un segment manquant sur une seconde plante **Figure 5A**. La reconstitution du génome complet se fait alors peu après l'entrée du virus dans le puceron. Les segments sont ingérés avec la sève de la plante infectée, puis internalisés et durablement stockés dans des vésicules cytoplasmiques de l'intestin moyen. Des acquisitions successives de différents segments par un même puceron vecteur, même à quelques jours d'intervalle, se cumulent dans ces vésicules et reconstituent ainsi le génome complet. D'autre part, deux pucerons ayant acquis le FBNSV chacun sur une plante dans laquelle un segment est absent, avec le segment absent différent pour chaque plante, peuvent reconstituer un génome viral complet en inoculant tous les deux une même plante receveuse **Figure 5B**. La possibilité d'acquisition et de transmission séparées des segments d'un virus multipartite réduit le coût engendré par le risque de perdre de l'information génétique lors du passage d'un hôte à un autre.

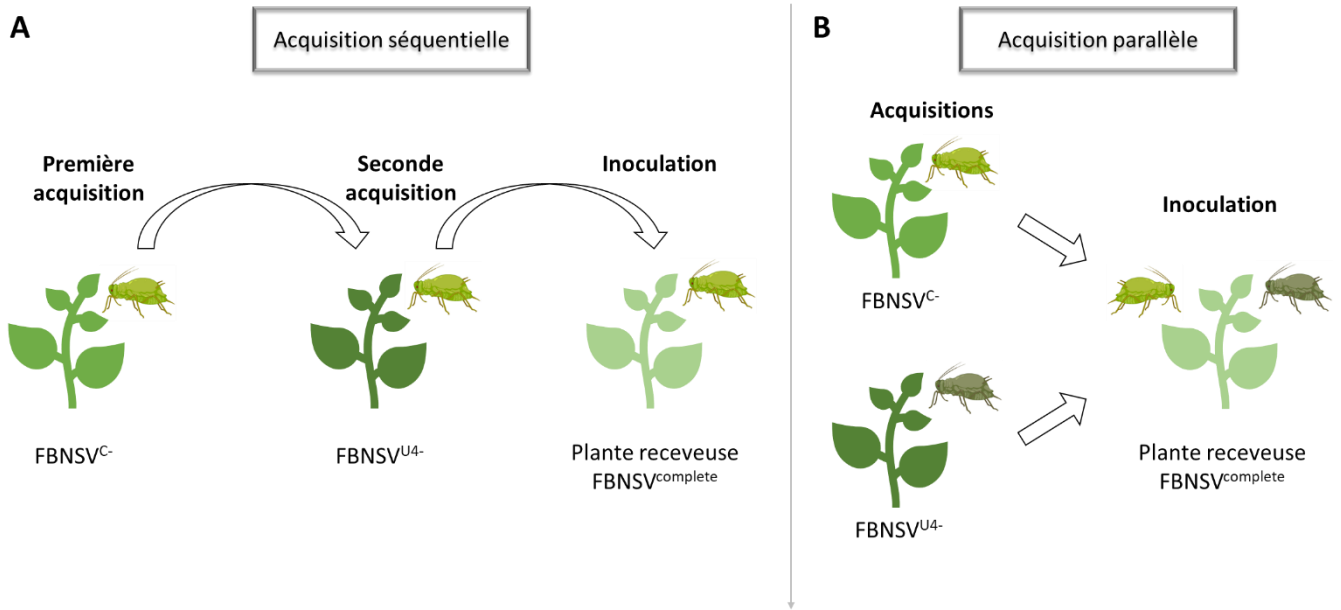


Figure 5 : Transmission séparée de deux segments génomiques du FBNSV. (A) Un puceron vecteur acquiert une partie du virus sur une plante dans laquelle le segment C est absent (FBNSV^{C-}). Il est ensuite transféré sur une seconde plante dans laquelle le segment U4 est absent (FBNSV^{U4-}), afin d'acquérir le segment manquant lors de la première inoculation. Enfin, il est transféré sur une dernière plante saine, sur laquelle il va inoculer le virus acquis sur les deux plantes précédentes. **(B)** deux pucerons acquièrent une partie du génome viral sur deux plantes différentes dans lesquelles il manque un segment FBNSV^{C-} ou FBNSV^{U4-}). Les deux pucerons sont ensuite transférés sur la même plante receveuse afin de reconstituer un génome viral complet.

Adapté de Di Mattia et al., 2022

D. Avantages potentiels du multipartitisme

Puisque le multipartitisme est apparu plusieurs fois et a été maintenu au cours de l'évolution (Lucía-Sanz and Manrubia, 2017), il pourrait présenter un ou plusieurs avantages sélectifs pour compenser le coût élevé généré par la perte d'une partie de l'information génétique. Des bénéfices potentiels ont été recensés par différents auteurs (Iranzo and Manrubia, 2012; Lucía-Sanz et al., 2018; Lucía-Sanz and Manrubia, 2017; Michalakis and Blanc, 2020; Sicard et al., 2016). La plupart des propositions détaillées ci-dessous, à l'exception de celle concernant la stabilité des particules, sont permises par la segmentation du génome viral et peuvent donc s'appliquer aussi bien aux virus segmentés que multipartites. Elles n'expliquent pas spécifiquement l'encapsidation séparée des segments. Les avantages proposés sont, soit conférés par la taille plus petite des molécules d'ADN ou d'ARN générée par la segmentation du génome (réplication et encapsidation plus rapides, taux de mutation plus faibles, stabilité des virions) ; soit, permis par la séparation de l'information génétique sur plusieurs molécules (réassortiments, expression).

1. Réplication

Les segments plus courts sont répliqués et encapsidés plus rapidement qu'une molécule plus longue, les rendant ainsi plus compétitifs lorsqu'ils sont dans une même cellule, si les quantités de réplicase et de protéine de capsid ne sont pas limitantes. Cette hypothèse est supportée par plusieurs modèles théoriques (Chao, 1991; Nee, 1987; Park et al., 2023; Pressing and Reaney, 1984). (Nee, 1987) soutient que cet avantage conféré aux segments courts suffirait à expliquer l'évolution vers la segmentation du génome, sans aucune fonction adaptative pour le virus. Selon son hypothèse, la segmentation correspondrait à un compromis entre l'avantage sélectif attribué par des acides nucléiques courts et la conservation d'un génome complet codant pour toutes les protéines essentielles. Par ailleurs, pour un taux de mutation par nucléotide donné, il est plus facile de produire des segments sans ou avec peu de mutations délétères quand la taille des segments est plus faible (Chao, 1991; Pressing and Reaney, 1984).

2. Stabilité des particules

Le foot-and-mouth disease virus (FMDV ; famille *Picornaviridae*) est un virus monopartite à ARN dont les infections successives de cultures cellulaires avec une forte MOI ont conduit à l'apparition spontanée d'une population virale composée de deux variants de taille inférieure au génome initial. Ces deux variants sont chacun délétés d'une partie différente du génome du FMDV et sont donc complémentaires. Cette population de variants, assimilée à un virus au génome segmenté (bipartite),

supplante le virus initial en condition de forte MOI (García-Arriaza et al., 2004). (Ojosnegros et al., 2011) ont étudié l'avantage sélectif de ce variant bipartite par rapport à une version de taille initiale issue de la recombinaison entre les deux génomes délétés. Les auteurs ont déterminé que la segmentation du génome n'entraîne d'avantage sélectif ni au niveau de la synthèse d'ARN, ni au niveau de la synthèse de protéines virales. Néanmoins, des tests d'infectivité suggèrent que la population au génome bipartite est plus infectieuse que la population au génome non segmenté. Cette différence d'infectivité est corrélée avec une inactivation plus lente des particules contenant les segments délétés lors d'un traitement thermique, indiquant une plus grande stabilité des particules contenant un génome de plus petite taille. Ainsi, la plus grande stabilité des particules contenant des génomes courts confère un avantage sélectif à la population bipartite par rapport à la population initiale non-segmentée.

Une étude récente par Park et al., 2023 discute ces résultats obtenus dans l'étude précédente. Les auteurs indiquent que le cas du FMDV doit être une exception, puisque l'augmentation de la stabilité des particules virales conférée par les génomes défectifs du FMDV implique que la capsid ne soit pas adaptée au virus initial monopartite. Park et al., 2023 s'appuient sur des données de la littérature indiquant que la capsid de certains virus à RNA est capable d'encapsider des fragments d'ARN plus longs que le fragment initial. A l'inverse, l'encapsidation de fragments inférieurs à la taille du génome initial ne peut se faire que si plusieurs fragments sont encapsidés ensemble (Cadena-Nava et al., 2012). Ainsi, pour la majorité des virus la stabilité des capsides contenant le génome entier doit être meilleure ou égale à la stabilité des particules défectives (Park et al., 2023).

3. Expression des gènes viraux

La répartition des différents gènes viraux sur plusieurs molécules permet le contrôle dissocié de l'expression des différents gènes viraux *via* une régulation indépendante de la transcription ou de la traduction pour chaque segment génomique (Yu et al., 2019). Par exemple, chez les virus de la famille des *Nanoviridae*, chez lesquels presque tous les segments génomiques ne contiennent qu'une seule ORF, plusieurs études montrent ou suggèrent une force différentielle des promoteurs en fonction du segment génomique (Bashir et al., 2022; Dugdale et al., 2000, 1998; Gallet et al., 2022; Shirasawa-Seo et al., 2005; Yu et al., 2019). Cependant, cet avantage est à nuancer car chez de nombreux virus, chaque gène possède son propre promoteur même s'il est porté par un segment codant plusieurs gènes. C'est par exemple le cas des *Begomivirus* bipartites (famille *Geminiviridae*), dont les gènes portés sur un même segment génomique peuvent être régulés indépendamment.

La répartition des gènes sur plusieurs segments génomiques pourrait également permettre de séparer l'expression de ces gènes dans le temps et l'espace en fonction de la localisation intra-cellulaire des segments génomiques (Harrison et al., 1976).

4. Réassortiments génétiques

A l'inverse de Nee qui propose que la sélection conduisant à la segmentation du génome s'applique au niveau du segment individuel, Chao, 1991 soutient que la sélection doit agir sur l'ensemble des segments constituant le génome viral. La segmentation du génome permet des réassortiments de segments entre différents virus coinfectant la même cellule. Ces échanges permettraient d'une part de compenser l'apparition de mutations délétères engendrées par des taux de mutations élevés. De cette manière, la segmentation du génome serait sélectionnée par la pression de mutation et constitueraient une alternative à la recombinaison (Chao, 1991). D'autre part, les échanges génétiques entre génomes viraux augmentent les opportunités d'échanges de matériel génétique et la possibilité d'apparition de nouvelles combinaisons de gènes (Fulton, 1980).

E. Conclusion sur les coûts et avantages du multipartitisme

Dans une simulation, Leeks et al., 2023 déterminent que ni une MOI élevée (section C.1), ni un avantage au niveau du groupe de segments (section D.4 par exemple) ne sont nécessaires à l'évolution du monopartitisme vers le multipartitisme (ou la segmentation du génome). Lors d'une infection virale, des virus défectifs, ayant perdu la capacité de produire l'intégralité des produits de gènes viraux peuvent apparaître. Dans une population virale composée d'un virus ancestral monopartite et de virus défectifs pouvant être complétés entre eux ou par le virus ancestral, le bénéfice intrinsèque conféré à chaque défectif parce qu'il est plus court (sections D.1 et D.2) ou l'avantage de la séparation de l'information génétique (section D.3) suffisent à l'évolution vers un multipartite ou un segmenté. Néanmoins, les auteurs soulignent que le multipartitisme peut ensuite permettre l'apparition de nouvelles fonctions adaptatives, qui ne sont pas possibles chez l'ancêtre monopartite, et qui confèrent un avantage sélectif au groupe de segments.

Enfin, la plupart de ces propositions concernent la segmentation du génome, mais pas spécifiquement l'encapsidation multiple, qui reste par conséquent très mal, voire pas du tout comprise.

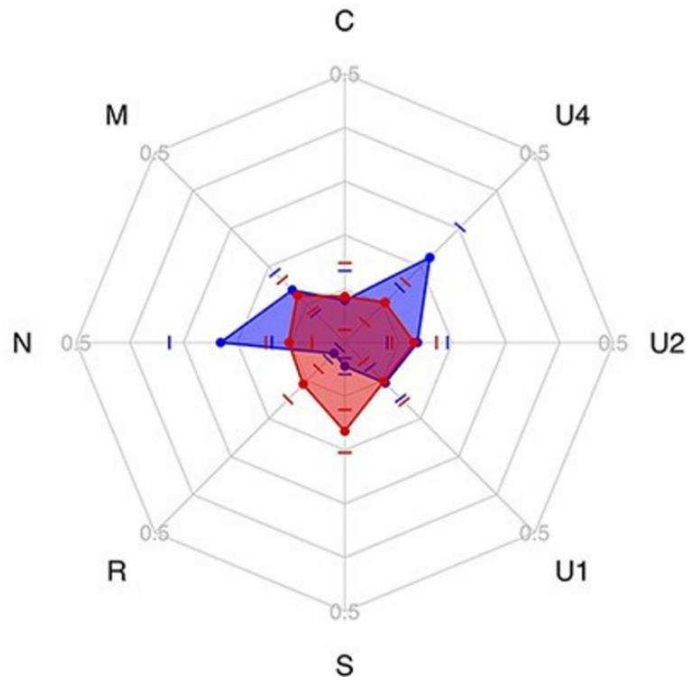


Figure 6 : Formule génomique du FBNSV dans *V. faba* (bleu) et *M. truncatula* (rouge). Les segments génomiques du FBNSV ont des fréquences relatives différentes au sein d'un même hôte. Le pattern d'accumulation varie entre les deux hôtes (pattern bleu et rouge). Les points correspondent à la médiane des fréquences pour chaque segment indiqué à l'extrémité de l'axe. Les barres représentent les écarts-types.

Gallet et al., 2022

II. La formule génomique

A. Définition

Une possibilité pour les virus segmentés et multipartites conférée par la nature fragmentée de leur génome est que les différents segments génomiques peuvent s'accumuler en quantités différentes dans l'hôte. Même, si certaines études avait préalablement remarqué des différences d'accumulation relatives des segments génomiques de certains virus segmentés ou multipartites (Feng et al., 2006; French and Ahlquist, 1988; Hajimorad et al., 1991), les recherches n'y ont pendant longtemps pas porté d'intérêt. En 2013, notre équipe a formellement quantifié l'accumulation des différents segments génomiques du FBNSV dans la plante hôte *V. faba* (Sicard et al., 2013). Les huit segments qui composent le génome du FBNSV s'accumulent à des fréquences relatives différentes dans des feuilles de *V. faba* infectées (**Figure 6**). Certains segments génomiques sont relativement peu accumulés (par exemple R et S), alors que d'autres sont très fréquents (par exemple N et U4). Bien que la variance inter-plantes pour chaque segment soit grande, la médiane est très reproductible entre différents lots de plantes infectées. Le pattern d'accumulation des différents segments, correspondant aux médianes des fréquences relatives de chacun d'eux, a été appelé « formule génomique ». Au cours de la progression de l'infection, la variabilité inter-plante diminue, indiquant que les fréquences dans chaque plante convergent vers des fréquences relatives spécifiques. Ces fréquences correspondant à une formule génomique d'équilibre (ou *set-point genome formula* ; SGF) qui n'évolue plus une fois que l'infection systémique est bien instaurée (Sicard et al., 2013). La formule génomique est similaire pour des extraits d'ADN viraux totaux (encapsidés et non encapsidés) et d'ADN encapsidés. Elle est aussi indépendante de la méthode d'inoculation (agro-infiltrations ou transmission par puceron) et des fréquences des segments dans l'*inoculum*. Consécutivement à cette étude, des formules génomiques ont été déterminées chez

- des virus multipartites à ADN infectant des plantes (Bashir et al., 2022; Guyot et al., 2022; Mansourpour et al., 2022; Yu et al., 2019) ou des insectes (Hu et al., 2016),
- des virus multipartites à ARN infectant des plantes (Wu et al., 2017) et
- des virus segmentés infectant des animaux (Diefenbacher et al., 2018; Moreau et al., 2020),

indiquant que la formule génomique doit être un phénomène commun pour des virus au génome segmenté.

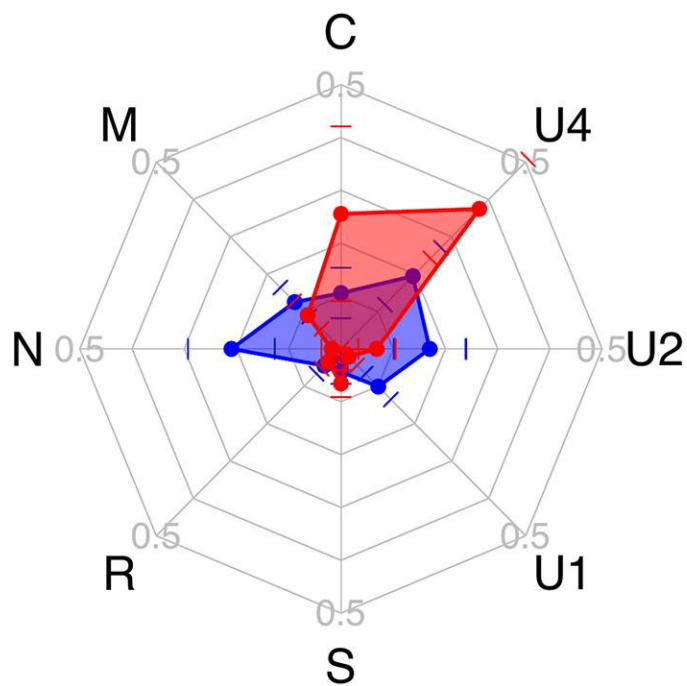


Figure 7 : Formules génomiques des nanovirus FBNYV (rouge) et FBNSV (bleu) dans *V. faba*. Les points correspondent à la médiane des fréquences pour chaque segment. Les barres représentent les écarts-types.

Mansourpour et al., 2022

La formule génomique d'un virus est spécifique au genre (Gallet et al., 2022; Moreau et al., 2020; Sicard et al., 2013; Wu et al., 2017) et à l'espèce (Torralba et al., 2023, en préparation) hôte. Par exemple, la formule génomique du FBNSV change lorsque le virus est transmis de *V. faba* à *Medicago truncatula* (Gallet et al., 2022; Sicard et al., 2013) (**Figure 6**). Par ailleurs, les formules génomiques du FBNSV dans deux variétés de *V. faba* (Shambat et Séville) sont similaires (résultats non publiés).

La formule génomique peut également varier drastiquement entre deux espèces virales. Le FBNSV et le FBNYV sont deux espèces de *Nanovirus* proches génétiquement (environ 75 % d'homologie de séquences) (Grigoras et al., 2014). Pourtant, leurs formules génomiques dans *V. faba* sont très différentes (Mansourpour et al., 2022) (**Figure 7**). Nous avons également observé des formules génomiques différentes pour trois isolats du FBNSV dans différents hôtes (Torralba et al., 2013, en préparation), indiquant que de petites variations de séquences virales peuvent impacter la formule génomique.

De prime abord, l'accumulation différentielle des segments paraît accentuer le coût généré par la segmentation du génome. En effet, le risque de perdre un segment est augmenté si certains sont peu représentés dans la population virale. Néanmoins, la formule génomique est présente chez plusieurs types viraux aux origines différentes, ce qui laisse penser qu'elle doit avoir un intérêt sélectif pour le virus.

B. Rôle fonctionnel de la formule génomique

Certaines observations de Sicard et al., 2013 suggèrent un lien causal entre la distance à la formule et l'augmentation de la charge virale totale et l'inhibition du développement de la plante causée par l'infection de *V. faba* par le FBNSV : (i) la distance à la SGF diminue en même temps que l'infection progresse, ce qui indique que dans chaque plante, la population virale progresse vers la SGF ; (ii) la charge virale totale augmente avec le temps ; (iii) l'accumulation virale augmente plus tôt dans les plantes où la distance à la formule minimale est atteinte le plus rapidement ; (iv) l'accumulation virale augmente plus rapidement dans des plantes dont le développement est stoppé tôt. La charge virale et l'intensité des symptômes semblent donc être des indicateurs de la fitness du virus. Ainsi, plus les fréquences relatives des segments génomiques se rapprochent de la SGF, plus la fitness du virus semble augmenter, suggérant une fonction adaptative de la formule génomique. Néanmoins, aucune corrélation entre charge virale totale et distance à la formule n'a été directement validée.

Le caractère adaptatif de la formule génomique pourrait être conférée par la possibilité de réguler l'expression des gènes viraux en modulant leur nombre de copies (Gutiérrez and Zwart, 2018; Michalakis and Blanc, 2020; Sicard et al., 2016, 2013; Zwart and Elena, 2020). Ce mécanisme, connu sous le nom de *gene copy number variation* (CNV), a été décrit pour les micro-organismes procaryotes et eucaryotes, ainsi que pour les grands virus à ADN double brin monopartite, pour lesquels il permet une adaptation rapide de l'expression des gènes dans de nouveaux environnements (Bayer et al., 2018; Elde et al., 2012; Tomanek et al., 2020).

L'hypothèse d'un mécanisme analogue au CNV chez les virus multipartites a été modélisée par Zwart and Elena, 2020. En étudiant les conditions pour lesquelles un virus bipartite pourrait surpasser un virus équivalent monopartite, les auteurs ont déterminé que plus les environnements sont hétérogènes et plus la nécessité de spécificité d'expression des gènes pour chaque environnement augmentent, plus le virus bipartite est apte à surpasser le monopartite. Ainsi, dans certaines conditions, le coût du multipartitisme peut être contrebalancé par l'avantage d'un ajustement rapide de l'expression des gènes, confortant l'hypothèse que la formule génomique pourrait jouer un rôle d'adaptation à des environnements changeants.

Expérimentalement, Gallet et al., 2022 ont déterminé que l'accumulation de chaque segment génomique du FBNSV dans *V. faba* et *M. truncatula* est positivement corrélée à l'accumulation des ARNs messagers correspondants (et pas à ceux des autres segments). Ces résultats montrent que des variations de la formule génomique peuvent moduler l'expression des gènes et confortent l'hypothèse d'un ajustement du nombre de copies de gènes afin d'adapter rapidement leur expression en réponse à un environnement variable.

C. Etablissement de la formule génomique

Gallet et al., 2022 montrent que le changement de formule génomique entre *V. faba* et *M. truncatula* n'est pas lié à une variation de la fréquence de mutation du génome du FBNSV pendant ce changement d'hôte. Les mécanismes mutation-indépendants régissant la formule génomique sont inconnus. Plusieurs hypothèses impliquant différents niveaux de régulation ont été proposées. La formule génomique pourrait résulter d'une régulation au niveau de chaque segment individuellement : par

exemple, des taux de réplication différents selon le segment (Sicard et al., 2013), une encapsidation différentielle ou une stabilité des particules virales variable en fonction du segment génomique.

Une autre possibilité est que la formule génomique résulte d'une sélection agissant au niveau du groupe de segments lors de l'infection systémique. Les goulots d'étranglement sévères (cf. section C.1) entraînent une diminution drastique et un échantillonnage aléatoire de génomes qui vont initier de nouveaux points d'infection. Ce phénomène génère une dérive de la formule génomique (Gutiérrez and Zwart, 2018; Zwart and Elena, 2020). Ainsi, lors des premiers stades de l'infection, des groupes de segments ayant des fréquences relatives différentes peuvent chacun initier des foyers d'infection différents. Ensuite, les groupes possédant les fréquences le plus optimales pour le système viral se propagent plus rapidement. Ils sont ainsi sélectionnés positivement au fil des cycles d'infection successifs, jusqu'à constituer la quasi-totalité de la population virale (Gutiérrez et Zwart, 2018 ; Sicard et al., 2016, 2013 ; Zwart et Elena, 2015). Cette hypothèse est étayée par des études sur les virus à ARN montrant que le très petit nombre de génomes qui infectent une nouvelle cellule peut générer des points d'infection avec des compositions génétiques différentes. Ceci augmente le potentiel de sélection entre les différents points d'infection au sein d'un même hôte. Les génomes adaptatifs sont alors sélectionnés, élargissent l'infection virale et finissent par représenter l'ensemble de la population (Miyashita et al., 2015 ; Miyashita et Kishino, 2010). De cette manière, la formule génomique résulterait de l'addition d'événements stochastiques engendrés par la distribution aléatoire de chaque segment génomique lors d'un nouvel événement d'infection, et d'une sélection de groupes de segments dont les fréquences sont les plus adaptées au contexte cellulaire, *i.e.* à la SGF (Gutiérrez and Zwart, 2018; Sicard et al., 2013; Zwart and Elena, 2020).

L'hypothèse de sélection de groupes de segments a été modélisée par Zwart and Elena, 2020 pour un virus bipartite. Elle montre que la convergence vers la SGF est très rapide et survient après quelques cycles d'infection cellulaire seulement. Cette étude met également en évidence que le temps de convergence vers la SGF dépend de la MOI. Avec une MOI faible, les variations de formule génomique entre les cellules sont faibles, mais il y a une forte sélection pour l'infectivité (les deux segments doivent être présents dans la cellule) qui permet des changements rapides. Une MOI intermédiaire diminue la force de sélection, mais génère de plus grosses variations de formules génomiques. Ainsi, des MOI faibles ou intermédiaires entraînent une convergence vers la SGF est rapide. A l'inverse, lorsque la MOI est forte, les taux d'infection sont élevés par rapport à la faible variabilité de la formule génomique, ce qui limite à la fois les variations et la sélection et donc la rapidité d'évolution de la formule génomique. De manière

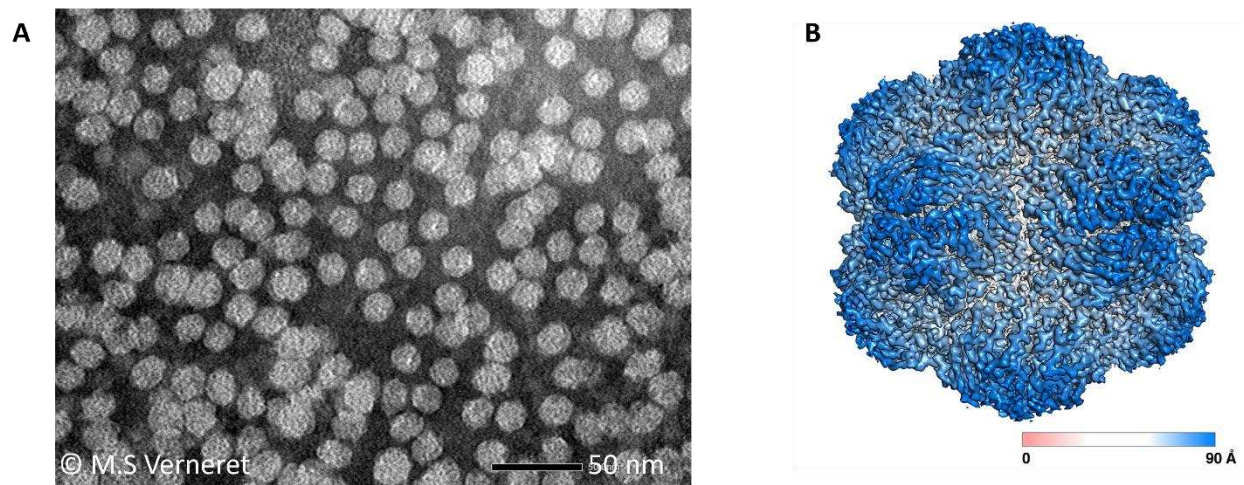


Figure 8 : Particules virales de *Nanovirus*. (A) Particules virales du PNYDV purifiées à partir de *Pisum sativum* L. et photographies au microscope électronique en coloration négative. (B) Isosurface de la densité de diffusion d'une particule de FBNSV (sharpenedmap) reconstruite à une résolution de 3,2 Å (EMD-10097). La surface est colorée en fonction de la distance par rapport au centre de la particule.

Panel B adapté de Trapani et al., 2023

cohérente avec cette modélisation, plusieurs études mettent en évidence que la MOI augmente au cours de l'infection (González-Jara et al., 2009; Gutiérrez et al., 2010; Tromas et al., 2014). Ainsi, dans les premiers stades de l'infection, une faible MOI pourrait entraîner des changements rapides de la formule génomique. Dans les stades les plus avancés, une forte MOI permettrait au virus d'infecter un grand nombre de cellules (Zwart and Elena, 2020).

III. Le FBNSV pour étudier la formule génomique

A. Généralités sur les *Nanoviridae*

1. Description

Les *Nanoviridae* sont une famille de phytovirus multipartites à ADN circulaire simple brin. Cette famille comporte deux genres : le genre *Babuvirus* (babuvirus) dont le génome est divisé en six segments génomiques ; et le genre *Nanovirus* (nanovirus) dont le génome est divisé en huit segments génomiques (Thomas et al., 2021). Les segments génomiques des *Nanoviridae*, leur structure et leur mode de réplication sont détaillés dans l'article de revue présenté dans le chapitre 2. Chaque segment est encapsidé individuellement dans une capside icosaédrique non enveloppée, d'environ 18-20 nm de diamètre et de symétrie T=1 (Lal et al., 2020b) (**Figure 8**). Chaque capside est constituée d'une soixantaine de protéines de capside et les particules virales sont *a priori* identiques quel que soit le segment encapsidé (Trapani et al., 2023).

Les *Nanoviridae* sont restreints aux cellules compagnes du phloème de leur hôte (Chu et al., 1995). De ce fait, leur transmission inter-hôte nécessite l'intervention d'un puceron vecteur (Lal et al., 2020b). Seulement deux espèces de pucerons ont été identifiés comme vecteurs de babuvirus : *Pentalonia nigronervosa* and *Micromyzus kalimpongensis*. Les nanovirus peuvent quant à eux être transmis par diverses espèces de pucerons, par exemple *Aphis craccivora*, *Acyrtosiphon pisum* et *Myzus persicae* (Lal et al., 2020b).

2. Gamme d'hôtes

Les virus du genre *Babuvirus* infectent uniquement des monocotylédones des familles *Musaceae* et *Zingiberaceae* (Lal et al., 2020b). Les *Nanovirus* ont historiquement été trouvés sur des *Fabaceae*. Plus récemment, de nouvelles études ont identifié de nouvelles familles de mono- et dicotylédones hôtes de

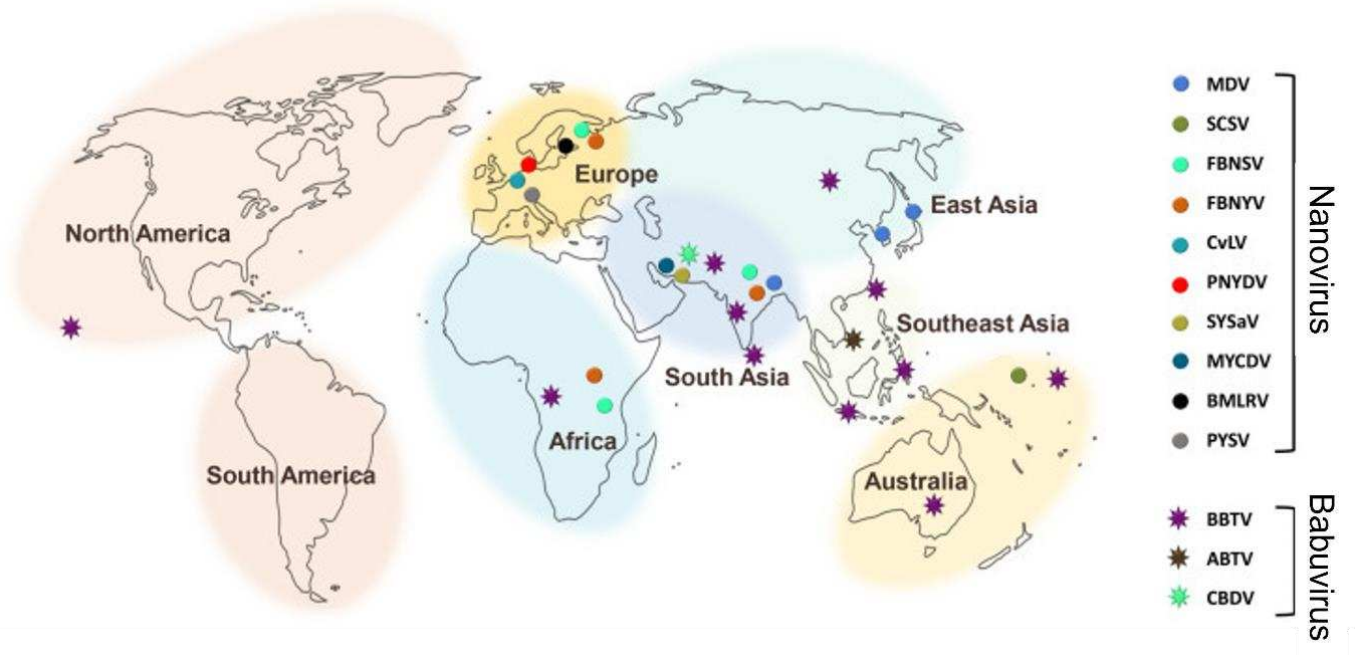


Figure 9 : Distribution géographique des *Nanoviridae*. Les étoiles représentent les espèces du genre *Babuvirus* et les cercles représentent celles du genre *Nanovirus*.

Adapté de Lal et al., 2020b

nanovirus, comme les *Apiaceae* (Hasanvand et al., 2021; Vetten et al., 2019), *Liliaceae* (Choi et al., 2019), *Solanaceae* et *Caricaceae* (Lal et al., 2020a, 2018) (synthétisé par (Lal et al., 2020b)). Le *Parsley severe stunt associated virus* (PSSaV), identifié sur le persil (*Petroselinum crispum*, *Apiaceae*), se distingue des autres espèces de nanovirus par le fait qu'un des huit segments génomiques, le segment U4, n'a pas été détecté chez les deux isolats identifiés (Hasanvand et al., 2021; Vetten et al., 2019).

3. Répartition géographique et importance économique

Les *Nanoviridae* sont présents en Europe, Asie, Afrique et Océanie. L'Amérique est épargnée, bien qu'une unique détection du BBTV a été observée à Hawaï (**Figure 9**). Alors que le BBTV est largement répandu sur les quatre continents, la plupart des *Nanoviridae* sont restreints à des régions relativement limitées du monde. Leur gamme d'hôtes étant assez restreinte, et focalisée sur les légumineuses, les *Nanoviridae* n'ont jusqu'alors pas été considérés comme une grande menace l'agriculture. Néanmoins, l'émergence de nouvelles espèces infectant un spectre d'hôtes plus large pourrait avoir des conséquences socio-économiques dévastatrices pour de nombreux pays (Lal et al., 2020b).

Le BBTV fait exception puisqu'il est responsable de l'une des plus anciennes maladies virales connues. La *banana bunchy top disease* est considérée comme la maladie virale la plus destructrice dans le monde de l'agriculture. Elle touche les cultures de bananiers et plantains, et entraîne des pertes de récoltes considérables. Sa large dispersion a été favorisée par des échanges de matériel végétal infecté (Qazi, 2016), donc principalement par l'activité humaine.

4. Les alpha-satellites

Les *Nanoviridae* peuvent être associés à des alpha-satellites (famille *Alphasatellitidae*) (Briddon et al., 2018). Ces derniers sont des molécules d'ADN circulaire simple-brin qui codent pour une unique protéine, une *replication initiator protein* (alpha-Rep), qui leur permet de se répliquer de manière autonome dans la cellule hôte (Horser et al., 2001; Timchenko et al., 1999). Néanmoins, les alpha-satellites dépendent d'un virus assistant pour l'encapsidation et les mouvements intra- et inter-hôte (Briddon et al., 2018).

Seulement trois études décrivent l'impact d'un alpha-satellite sur son virus assistant de la famille des *Nanoviridae*. Timchenko et al., 2006 ont étudié l'influence d'un alpha-satellite (le subterranean clover stunt alphasatellite 1 ; SCSA 1) sur un isolat égyptien (EV1-93) du nanovirus *Faba bean necrotic yellows virus* (FBNYV). Ce virus et cet alpha-satellite avaient été détectés dans le même échantillon prélevé dans la nature. La présence de l'alpha-satellite diminue les taux d'infection par agro-inoculation, ainsi que la sévérité des symptômes du FBNYV, suggérant une interférence entre le satellite et la multiplication du

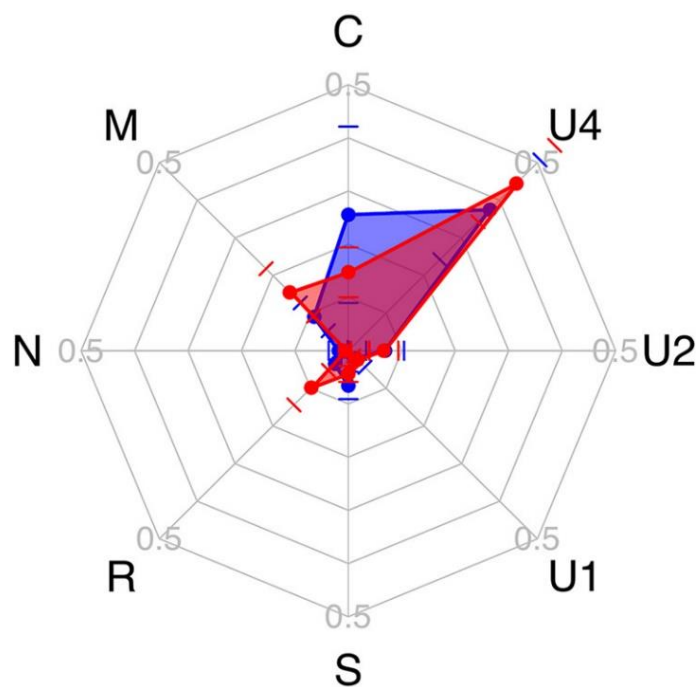


Figure 10 : Formule génomique du FBNYV dans *V. faba* en présence (rouge) ou en absence (bleu) de l'alpha-satellite SCSA1. Les points correspondent à la médiane des fréquences pour chaque segment. Les barres représentent les écart-types. Les fréquences relatives des segments C, M, N, R, S, et U4 sont significativement impactées par la présence de l'alpha-satellite.

Mansourpour et al., 2022

FBNYV, probablement due à une compétition pour les facteurs nécessaires à l'accomplissement du cycle viral (réplication, encapsidation et mouvement systémique) (Timchenko et al., 2006). Une étude postérieure étudie l'impact d'une co-infection de ce même alpha-satellite avec un isolat espagnol du FBNYV (isolat ES ; Mu29). A l'inverse de l'étude précédente, la présence de l'alpha-satellite n'affecte ni la sévérité des symptômes, ni l'accumulation du nanovirus dans *V. faba*, mais augmente le taux de transmission du FBNYV par son insecte vecteur (Mansourpour et al., 2022). Enfin, chez les babuvirus, l'association d'un alpha-satellite avec le BBTV dans des bananiers diminue l'accumulation et la transmission du BBTV par puceron (Guyot et al., 2022). Ainsi, l'impact d'un alpha-satellite doit dépendre du tryptique plante-virus-satellite.

Par ailleurs, les études de Mansourpour et al., 2022 et Guyot et al., 2022 montrent que la présence de l'alpha-satellite modifie l'accumulation relative des segments génomiques, *i.e.* la formule génomique, du virus assistant (**Figure 10**).

B. Notre système d'étude FBNSV – *Vicia faba*

Le *Faba bean necrotic stunt virus* (FBNSV) appartient au genre *Nanovirus*. Son génome est donc composé de huit segments : C, M, N, R, S, U1, U2 et U4. Nous disposons d'un clone infectieux de l'isolat JKI-2000 du FBNSV produit par Grigoras et al., 2009. Ce clone consiste en huit plasmides binaires (pBIN-19), chacun contenant une répétition en tandem d'un des segments. Chaque plasmide a été utilisé pour transformer une colonie d'*Agrobacterium tumefaciens* (souche COR308), produisant ainsi huit clones bactériens chacun capable de transférer un des segments du FBNSV dans la plante hôte *via* une agro-infiltration.

Durant ma thèse, j'ai majoritairement utilisé la fève (*V. faba*) comme hôte. Dans nos conditions de culture, les premiers symptômes de l'infection de *V. faba* par le FBNSV apparaissent aux environs de sept à dix jours après l'inoculation. D'abord, les feuilles les plus jeunes s'enroulent sur elles-mêmes. Par la suite, la croissance de la plante est stoppée et les feuilles de l'apex déjà enroulées s'épaississent. Dans un stade plus avancé de la maladie, les feuilles jaunissent et fissent par se nécroser (**Figure 11**).

Les huit segments génomiques du FBNSV ont systématiquement été détectés dans des échantillons prélevés dans la nature. Néanmoins, en conditions expérimentales, il arrive que certains segments dits « dispensables » soient perdus. En omettant d'inoculer un à un les différents segments du

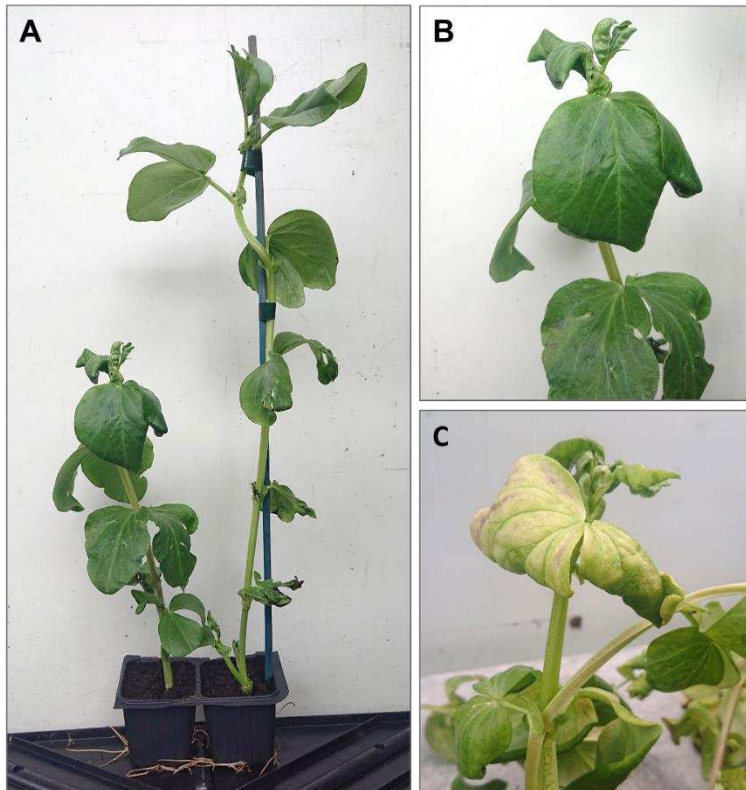


Figure 11 : Symptômes causés par une infection par le FBNSV chez *V. faba*. (A) L'infection par le FBNSV entraîne un arrêt de la croissance et conduit donc à un nanisme des plantes infectées (droite) par rapport aux plantes saines (gauche). **(B)** La maladie provoque un enroulement des feuilles des étages supérieurs. **(C)** Dans un stade plus avancé de la maladie, les feuilles jaunissent et nécrosent.

FBNSV, Grigoras et al., 2018 ont démontré que les segments C, N, U1, U2 et U4 sont dispensables à l'infection de la fève (*Vicia faba*). Une infection en omettant à la fois U1 et U2 est également possible. Cependant, en fonction du ou des segments absents, le phénotype de l'infection peut être plus ou moins impacté. En revanche, il ne peut pas y avoir d'infection si un des segments M, R ou S est absent.

C. Les protéines codées par le FBNSV

Chacun des huit segments génomiques du FBNSV ne contient qu'un seul cadre de lecture (ORF) orienté dans le sens du virion et permet donc la production d'une seule protéine virale (Thomas et al., 2021). La fonction et le poids moléculaire des protéines portées par les différents segments sont résumés dans le **Tableau 1**.

Le segment C code pour la protéine Clink, pour *cell cycle link*, capable d'interagir avec des régulateurs du cycle cellulaire (Aronson et al., 2000). L'interaction de Clink avec des protéines de la famille des rétinoblastomes (pRB), *via* le motif LXCXE, stimule la réplication virale (Aronson et al., 2000; Wanitchakorn et al., 2000). Les pRBs sont des répresseurs de l'expression de gènes nécessaires à la progression du cycle cellulaire vers la phase S de réplication de l'ADN. En interagissant avec ces pRBs, Clink lève cette répression.

La protéine codée par le segment M comporte un domaine hydrophobe caractéristique des protéines transmembranaires (Burns et al., 1995; Katul et al., 1997). De manière cohérente, l'expression d'un transgène portant l'ORF du segment M du BBTV fusionnée à la séquence *GFP*, dans des cellules embryogéniques de bananier, permet de localiser la protéine en périphérie de ces cellules associée à la membrane plasmique (en absence du virus) (Wanitchakorn et al., 2000). Ce type de domaine a été identifié sur la protéine de mouvement (MP) des *Geminiviridae*. Par analogie, il a été établi que le segment M des *Nanoviridae* code pour une MP nécessaire aux mouvements de cellule à cellule et longue-distance.

De la même manière, le segment N code pour une protéine nommée *Nuclear Shuttle Protein* (NSP) en raison d'analogies de localisation cellulaire avec à la protéine NSP codée par certains *Geminiviridae*, (Sanderfoot and Lazarowitz, 1995; Wanitchakorn et al., 2000). Chez les géminivirus, NSP sert de navette pour le transport de l'ADN viral du noyau vers le cytoplasme (Sanderfoot and Lazarowitz, 1995). Chez les *Nanoviridae*, la fonction de NSP est moins claire. Par analogie aux *Geminiviridae*, il a été proposé que la NSP du babuvirus BBTV aide au transport des protéines ou ADN viraux hors du noyau des cellules végétales

Tableau 1 : Caractéristiques des protéines codées par les segments du FBNSV

Segment génomique	Protéine	Fonction	Références
C	Clink	Régulation du cycle cellulaire	Aronson <i>et al.</i> , 2000
M	Movement protein (MP)	Mouvements de cellule à cellule et longue distance	Wanitchakorn <i>et al.</i> , 1997
N	Nuclear shuttle protein (NSP)	Transmission par puceron vecteur	Grigoras <i>et al.</i> , 2018 ; Di Mattia <i>et al.</i> , 2022
R	Replication-associated protein (Rep)	Réplication	Timchenko <i>et al.</i> , 1999, 2000
S	Capsid protein (CP)	Encapsidation, mouvement longue-distance	Katul <i>et al.</i> , 1997 ; Wanitchakorn <i>et al.</i> , 1997
U1	U1	Inconnue	-
U2	U2	Inconnue	-
U4	U4	Inconnue	-

(Ji et al., 2019). Néanmoins, l'absence du segment N n'a aucun impact sur l'infection de *V. faba* par le FBNSV, mais abolit la transmission par le puceron vecteur (Grigoras et al., 2018). Il est désormais établi que la NSP est un « facteur assistant » de la transmission du FBNSV par son puceron vecteur (Di Mattia et al., 2020b, 2023; Grigoras et al., 2018). Le mécanisme de transmission du FBNSV, avec l'intervention de la NSP, est détaillé dans la section suivante.

Le segment R code pour la protéine Rep (ou M-Rep), pour Replication-associated protein, requise pour la réplication de tous les segments génomiques du virus (Timchenko et al., 2000, 1999). Son fonctionnement est détaillé dans l'article de revue présenté dans le chapitre 2.

Le segment S code pour la protéine de capsid (CP) virale (Katul et al., 1997; Wanitchakorn et al., 1997). L'assemblage de 60 copies de cette protéine constitue la capsid dont les caractéristiques ont été précédemment présentées dans la section II.A.1. L'introduction de mutations empêchant l'assemblage des copies de la CP du FBNSV entre elles abolit l'infection systémique de la plante, suggérant que les mouvements longue-distance du FBNSV dans la plante se font sous la forme de virions (Trapani et al., 2023).

Enfin, les segments U1, U2 et U4 codent pour des protéines aux fonctions encore inconnues. Quelques caractéristiques de ces protéines ont été identifiées. La protéine U1 possède un motif de signal de transport vers le noyau (NLS) putatif et son extrémité C-terminale présente des similarités de séquence avec une protéine promotrice de la réplication de certains géminivirus (REn) (Krenz et al., 2017). U4 possède un domaine transmembranaire, comme MP. L'expression de transgènes contenant la séquence de U1, U2 ou U4 du PNYDV fusionnée à la GFP et leur localisation dans des feuilles de *N. benthamiana* peut apporter des indices quant au rôle de ces protéines (Krenz et al., 2017). La protéine U1 fusionnée à la GFP (U1::GFP) est présente uniquement sous forme de cercles dans le noyau. L'introduction de mutations dans le motif NLS altère en partie la localisation nucléaire. U2::GFP est localisée dans le noyau et sous forme de petits points en périphérie des cellules. Enfin, U4 interagit avec MP et les complexes sont localisés en périphérie des cellules et au niveau du réticulum endoplasmique cortical. Néanmoins, ces résultats doivent être interprétés de manière précautionneuse. D'une part, les fusions avec les protéines marqueurs (GFP) peuvent affecter la fonctionnalité et la localisation des protéines. D'autre part, la localisation de chaque protéine a été déterminée indépendamment de la présence des autres protéines virales (sauf U4 et MP) et en absence du virus lui-même.

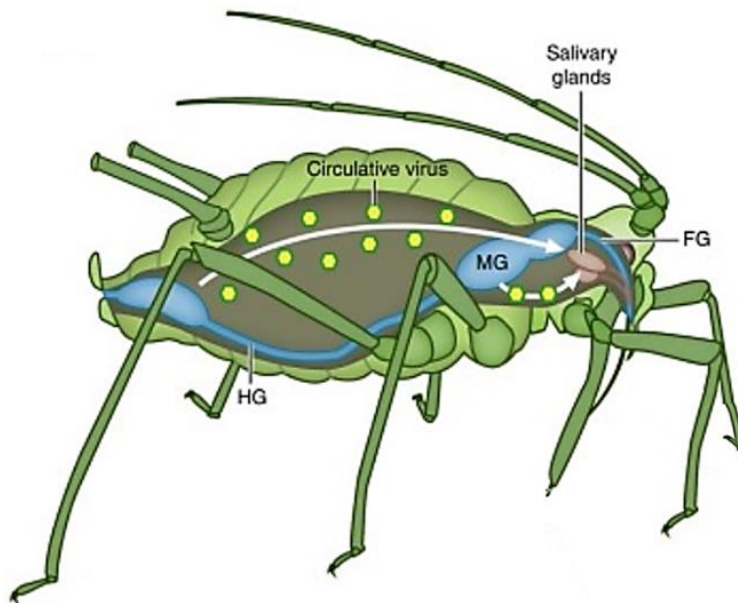


Figure 12 : Trajet d'un virus transmis selon le mode circulant non-multipliant dans son puceron vecteur. Les flèches blanches représentent le cycle de circulation des virus (hexagones jaunes) à l'intérieur du corps du puceron. Le virus traverse l'épithélium intestinal puis transite par l'hémolymphe vers les glandes salivaires. FG : intestin antérieur (*foregut*) ; MG : intestine moyen (*midgut*) ; HG ; intestin postérieur (*hindgut*)

Adapté de Blanc et al., 2014

D. Transmission du FBNSV

Comme tous les *Nanoviridae*, le FBNSV est transmis uniquement par pucerons et selon le mode circulant non-multipliant (Lal et al., 2020b). Après ingestion, le virus parcourt le corps du puceron vecteur en traversant d'abord la barrière intestinale, puis transite *via* l'hémolymphe jusqu'aux glandes salivaires et enfin la salive, sans jamais se répliquer. Il y a ainsi un temps de latence entre l'acquisition du virus et le moment où il pourra être inoculé *via* la salive du puceron (Blanc et al., 2014; Dietzgen et al., 2016) (**Figure 12**).

Les mécanismes moléculaires régissant la transmission du FBNSV ont été particulièrement étudiés par notre équipe. Il a été mis en évidence que NSP est nécessaire à l'internalisation du virus et son accumulation dans les cellules intestinales du puceron *Acyrtosiphon pisum*. De plus, la protéine NSP colocalise avec l'ADN viral et la protéine de capsid dans les cellules intestinales (Di Mattia et al., 2020b). Ces résultats suggèrent que NSP pourrait permettre l'internalisation du virus dans l'intestin en se fixant d'une part à un récepteur cellulaire au niveau de l'intestin antérieur et d'autre part à la particule virale. Etrangement, en réalisant des acquisitions séquentielles de différents segments du FBNSV, Di Mattia et al., 2022 ont mis en évidence que le virus peut être transmis même lorsque NSP est acquise jusqu'à trois jours après le virus. Pourtant, le mode d'action des facteurs assistants connu, et considéré universel jusqu'à récemment, implique que ceux-ci soient acquis avant ou en même temps que les particules virales. De cette manière, ils permettent l'attachement des particules virales au niveau de site de rétention dans le vecteur. La séquence d'acquisition inverse pouvant aboutir à une transmission n'a été démontrée pour aucun autre virus, les virions étant systématiquement perdus avec le transit intestinal en l'absence du facteur assistant (Di Mattia et al., 2023). Les résultats de transmission positive obtenus avec les deux séquences d'acquisition de la protéine NSP et des particules virales du FBNSV révèlent un mode d'action jusqu'à maintenant unique qui reste à élucider (Di Mattia et al., 2023).

Par ailleurs, Sicard et al., 2015 ont mis en évidence que les fréquences relatives de certains segments génomiques du FBNSV dans des pucerons, vecteurs ou non, sont différentes de la formule génomique dans la plante source (**Figure 13**). Notamment, la fréquence du segment N diminue et celle du segment U2 augmente fortement dans le puceron par rapport à la plante source. La formule génomique dans le puceron est similaire pour trois espèces testées (*Acyrtosiphon pisum*, *Aphis craccivora*, et *Myzus persicae*). Les changements de fréquences se font dès l'entrée dans l'intestin du puceron et leur cause est à ce jour un mystère.

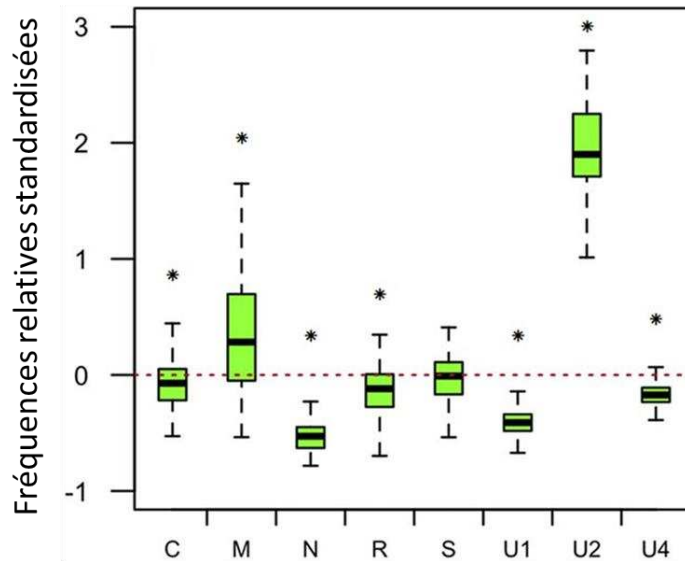


Figure 13 : Fréquences relatives des segments génomiques du FBNSV dans des pucerons *A. pisum* par rapport à la formule des plantes sur lesquelles ils se sont nourris. La fréquence relative standardisée correspond à la fréquence relative d'un segment dans les pucerons par rapport à leur fréquence dans les plantes sources correspondantes. Pour chaque diagramme en boîte, le bas et le haut de la boîte correspondent aux premier et troisième quartiles, la ligne horizontale à l'intérieur des diagrammes en boîte représente la valeur médiane de la distribution, et les moustaches délimitent 1,5 fois la distance entre les premier et troisième quartiles de la distribution. Les astérisques indiquent des différences significatives dans les fréquences d'un segment dans la plante source et dans les pucerons correspondants.

Adapté de Sicard et al., 2015

IV. Problématique et objectifs de thèse

La formule génomique paraît être une caractéristique commune aux virus segmentés et multipartites, et pourrait jouer un rôle fondamental dans leur fonctionnement. Néanmoins, les phénomènes qui la régulent sont inconnus. Ainsi, l'objectif de cette thèse est d'élucider les mécanismes régissant la mise en place et la régulation de la formule génomique. Nous avons d'abord exploré les données de la littérature sur la réplication des nanovirus, afin d'y déceler des indices sur une éventuelle réplication différentielle des segments génomiques. Nous avons ensuite testé expérimentalement les deux niveaux de régulation proposés dans la section II.C : une régulation de l'accumulation de chaque segment de manière individuelle ou une sélection à l'échelle du groupe de segments. Dans une dernière partie, nous nous sommes intéressés aux facteurs déterminant la formule génomique. Puisque la formule génomique est spécifique du génotype de l'hôte (Mansourpour et al., 2022; Moreau et al., 2020; Sicard et al., 2013; Wu et al., 2017), et qu'elle pourrait de ce fait avoir une fonction d'adaptation à l'hôte, nous avons recherché si des variations de la physiologie de la plante hôte influencent l'accumulation relative des segments génomiques du FBNSV.

Chapitre 1 : Etude bibliographique : la réplication des nanovirus et sa potentielle implication dans l'établissement de la formule génomique

I. Contexte

Plusieurs hypothèses sur les mécanismes de mise en place de la formule génomique sont exposées dans la littérature (section II.C de l'introduction). L'une d'entre elles propose que la formule génomique résulte d'une réplication différentielle des segments génomiques (Sicard et al., 2013). Nous avons exploré cette hypothèse en collectant les données sur la réplication des *Nanoviridae* existantes dans la littérature. Nous avons constaté que les recherches sur la réplication des nanovirus sont peu nombreuses et très souvent déduites des connaissances sur les *Geminiviridae*. Les *Geminiviridae* sont l'unique autre famille de virus à ssDNA infectant les plantes. Les *Nanoviridae* et *Geminiviridae* présentent des similarités au niveau de leur organisation génomique suggérant des modes de réplication similaires. *A contrario* des nanovirus, les mécanismes de réplication de géminivirus a été largement étudiée et sont bien compris.

Ainsi, nous avons rédigé une synthèse bibliographique faisant l'état des lieux des connaissances sur la réplication de ces deux familles de virus. A partir des informations récoltées, nous discutons d'un rôle éventuel de la réplication dans l'instauration de la formule génomique chez les nanovirus. Nous proposons également un potentiel mécanisme analogue, permettant la régulation de l'expression des gènes, chez les *Geminiviridae*. Ce travail a fait l'objet d'un article de revue qui a été accepté pour publication dans mBio.

II. Article de revue : Replication mechanisms of circular ssDNA plant viruses and their potential implication in viral gene expression regulation

Mélia Bonnamy^{1,2}, Stéphane Blanc^{1*}, Yannis Michalakis^{2*}

¹ PHIM, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

² MIVEGEC, CNRS, IRD, Univ Montpellier, Montpellier, France

* These authors contributed equally to this work

Corresponding authors:

Stéphane Blanc (stephane.blanc@inrae.fr), Yannis Michalakis (yannis.michalakis@ird.fr)

Abstract

The replication of members of the two circular single-stranded DNA (ssDNA) virus families *Geminiviridae* and *Nanoviridae*, the only ssDNA viruses infecting plants, is believed to be processed by Rolling-Circle Replication (RCR) and Recombination-Dependent Replication (RDR) mechanisms. RCR is a ubiquitous replication mode for circular ssDNA viruses and involves a virus-encoded Replication-associated protein (Rep) which fulfills multiple functions in the replication mechanism. Two key genomic elements have been identified for RCR in *Geminiviridae* and *Nanoviridae*: (i) short iterative sequences called iterons which determine the specific recognition of the viral DNA by the Rep; and (ii) a sequence enabling the formation of a stem-loop structure which contains a conserved motif and constitutes the origin of replication. In addition, studies in *Geminiviridae* provided evidence for a second replication mode, RDR, which has also been documented in some double-stranded DNA viruses. Here, we provide a synthesis of the current understanding of the two presumed replication modes of *Geminiviridae* and *Nanoviridae*, we identify knowledge gaps and discuss the possibility that these replication mechanisms could regulate viral gene expression through modulation of gene copy number.

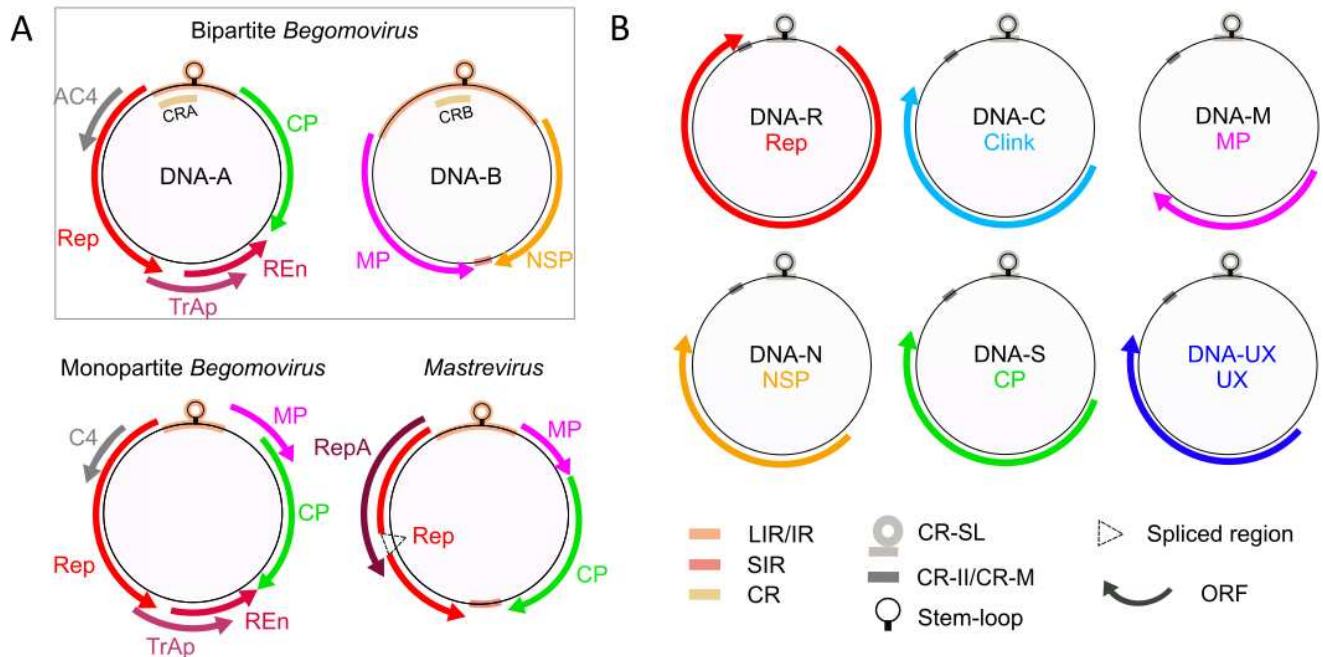


Fig. 1: Genome organization of the families *Geminiviridae* and *Nanoviridae*

A- Genome organization in the genera *Begomovirus* and *Mastrevirus* belonging to the family *Geminiviridae*. *Geminiviridae* are classified into 14 genera, which can be found in (7). All *Geminiviridae* genera have a genome constituted by a single DNA molecule (black circles) except for the genus *Begomovirus* which contains species that can be either monopartite or bipartite with complementary DNA-A and DNA-B. Solely the genomes of begomoviruses and mastreviruses are shown here because most of the experimental work we review has been obtained with these virus models.

B- Genome organization in the family *Nanoviridae*. The name of each component is indicated within the corresponding circle. The five genomic components DNA-R, C, M, N and S are identified for species of both genera *Nanovirus* and *Babuvirus*. DNA-UX represents components for which no function has so far been determined. The genus *Nanovirus* possesses three components of unknown function named DNA-U1, -U2 and -U4. The genus *Babuvirus* possesses only one component of unknown function named DNA-U3.

The colored arrows around the circles represent the open reading frames (ORF). For *Geminiviridae* genomes, arrow orientation reflects the position on the viral or complementary strand and alternatively spliced regions are indicated by dotted triangles. Intergenic and conserved non-coding regions are represented by rectangles inside or overlapping the black circles, and their identity is indicated below the panel. At the top of each DNA circle, a highly conserved stem-loop structure is represented. Rep: replication associated protein; repA and REn: replication enhancer proteins; TrAp: transcriptional activator protein; CP: capsid protein; MP: movement protein; NSP: nuclear shuttle protein; AC4/C4: protein with multiple functions; U1, U2, U3 and U4 (UXs): proteins with unknown functions; LIR: long intergenic region; SIR: small intergenic region; CRA and CRB: common regions of DNA-A and DNA-B corresponding to the highly conserved 200-nucleotide stretch within the large intergenic region of bipartite begomoviruses ; CR-SL: common region-stem loop; CR-II: second common region (*Nanovirus*); CR-M: major common region (*Babuvirus*). Adapted from (7, 19).

Introduction

Replication is a critical step in the life cycle of viruses as it primarily permits to generate progeny that will contribute to virus spread. One important category of viral replication is that of the single-stranded DNA (ssDNA) viruses, which does not involve virus encoded polymerases, but viral replication-associated proteins referred as Rep (1). DNA replication controlled by Rep proteins has been reported for bacterial plasmids, and an immense paraphyletic group of small ssDNA viruses infecting prokaryotes and all sorts of eucaryotes, including invertebrates, vertebrates and plants (2). Whether and how the replication process can fulfill additional functions, other than progeny production, is an issue that we illustrate through some speculations resulting from the survey of the replication mechanisms in the two families *Geminiviridae* and *Nanoviridae* that encompass all currently known ssDNA viruses infecting plants.

Members of the family *Geminiviridae* (geminiviruses) have a highly diversified host range, infecting both monocotyledons and dicotyledons. They are responsible for considerable losses on crops of major socio-economic interest to many countries, particularly in tropical and sub-tropical zones, such as for example maize, wheat, sugar cane, cassava, cucurbits, sweet potatoes, pepper, tomato and cotton. Their worldwide dispersal has been facilitated by human activities and mobility, agricultural intensification, and global warming expanding the geographical range of their vectors (3). Epidemics of new emerging geminiviruses are frequent (4). Recently, the emergence of new species, infecting perennial crops of major socio-economic interest also in temperate regions, including grapes, has drawn particular attention (5, 6). The appearance of new species is due in part to their high rate of mutation, recombination and reassortment, making this virus family one of the most diverse (3). It is therefore of paramount interest to understand the replication processes generating such a genetic diversity.

The family *Geminiviridae* is composed of 14 genera (Fig. 1A; (7)). Most geminiviruses, have their genetic information distributed over one single-stranded circular DNA molecule. However, members of genus *Begomovirus* can be mono- or bi-partite with their genome constituted either by one single DNA circle or by two complementary circles named DNA-A and DNA-B. For both mono- and bipartite species, the virus particle is a geminate structure resulting from the fusion of two icosahedral “subunits”, and each geminate particle contains a single DNA circle of comparable size ranging from 2.5 to 3.2 kb (8, 9). DNA-A and DNA-B of bipartite begomoviruses are therefore encapsidated in two distinct geminate virus particles. Each circular ssDNA molecule is composed of coding and non-coding regions (Fig. 1A). Coding regions are transcribed bidirectionally, and the number of open reading frames (ORF) differs according to the geminivirus genus (7, 10). The number and the name of the intergenic non-coding regions also vary

depending on the genus (Fig. 1A; (7)). Specifically, for the bipartite begomoviruses, the non-coding Intergenic Region (IR) contains a sequence of about 200 nucleotides defined as the Conserved Region (CR), that is identical between DNA-A and DNA-B (11). In all genera, the non-coding-regions contain regulatory sequences for replication and expression.

The family *Nanoviridae* (nanovirids), with their small number of identified species and restricted host range, have long been considered as a minor threat to agriculture (12). As a result, they have received little attention from the scientific community, whose research has tended to focus on geminiviruses. The *Banana bunchy top virus* (BBTV) is an exception, however, as it is responsible for banana bunchy top disease, one of the oldest known viral diseases of plants and considered to be the most destructive viral disease for the world agriculture (13). In recent years, epidemics due to the emergence of new nanoviruses have broadened the host spectrum, and thus the threat, to crops of prime importance for many countries, and above all numerous legume crops (12).

The family is thus far composed of solely two genera *Nanovirus* and *Babuvirus* (Fig. 1B). It is characterized by a multipartite genome architecture, where the genetic information is borne by eight (*Nanovirus*) or six (*Babuvirus*) ssDNA circular molecules called genome segments or components. Each component is individually encapsidated, all are about 1 kb long, and all encode a single protein and comprise a non-coding regulatory region. In all cases the ORFs are unique and transcribed unidirectionally in the virion-sense (14–17). The only known exception to this rule is for the BBTV DNA-R, where a second ORF coding for a small protein with unknown function has been identified nested within the Rep ORF (18). Non-coding regions contain two conserved sequences respectively called Common Region Stem-Loop (CR-SL) and Second Common Region (CR-M for *Babuvirus* or CR-II for *Nanovirus*; Fig. 1B; (19)). As for geminiviruses, the non-coding regions contain regulatory sequences for replication and expression.

While the replication of geminiviruses has been extensively studied, the replication of nanovirids has been little explored. Because these two families share some genomic similarities, most of the available information concerning nanovirids replication is either assumed by analogy or has been inferred by homology with geminiviruses. It is established that both geminiviruses and nanovirids genomes are replicated according to the Rolling-Circle Replication (RCR) (20, 21), a mechanism that has also been described for the replication of all the other eukaryotic circular ssDNA viruses (22), some bacterial plasmids (23) and some single- and double-stranded DNA phages (24). A series of studies conducted on geminiviruses suggested the existence of another mode of replication, also reported in some phages, the Recombination-Dependent Replication (RDR). Here, we survey the current knowledge on the replication

Table 1: Survey of RCR studies in *Geminiviridae* and *Nanoviridae*

Step	Special feature	<i>Geminiviridae</i>			<i>Nanoviridae</i>		
		Species	Genus	Reference	Species	Genus	Reference
Complementary-strand synthesis	DNA primers	MSV	<i>Mastrevirus</i>	(50, 52)	BBTV	<i>Babuvirus</i>	(90)
		DSV		(51)			
		CSMV		(53)			
		WDV		(54)			
		TobYDV		(55)			
	RNA primers	ACMV	<i>Begomovirus</i> (bipartite)	(49)			
Involvement of a host primase	TYLCV	<i>Begomovirus</i> (monopartite)	(56)				
	ToLCNDV	<i>Begomovirus</i> (bipartite)	(57)				
Complementary-strand synthesis involves host polymerase α	TYLCV	<i>Begomovirus</i> (monopartite)	(59)				
Viral DNA recognition	Identification of iteron sequences on the right side of the stem-loop sequence only	-	-	(67)			
		TGMV	<i>Begomovirus</i> (bipartite)	(63, 64, 74)			
		BGMV	<i>Begomovirus</i> (bipartite)	(74)			
		SqLCV	<i>Begomovirus</i> (bipartite)	(64)			
		BCTV	<i>Curtovirus</i>	(75, 131)			
		TYLCV	<i>Begomovirus</i> (monopartite)	(132, 133)			
		ToLCV-Nde	<i>Begomovirus</i> (bipartite)	(76)			
		ToMoTV	<i>Begomovirus</i> (bipartite)	(134)			
	Identification of iteron sequences on both sides of the stem-loop sequence	WDV	<i>Mastrevirus</i>	(71, 72)	FBNYV	<i>Nanovirus</i>	(95, 96)
		MSV	<i>Mastrevirus</i>	(73)	MDV	<i>Nanovirus</i>	(96)
		MYMIV	<i>Begomovirus</i> (bipartite)	(70)	SCSV	<i>Nanovirus</i>	(96)
					BBTV	<i>Babuvirus</i>	(97, 99)
	Rep recognizing the common region/iterons of the viral genome in a sequence-specific manner	TGMV	<i>Begomovirus</i> (bipartite)	(63–66, 68, 69, 74, 135)	ABTV	<i>Babuvirus</i>	(99)
		BGMV	<i>Begomovirus</i> (bipartite)	(74, 135)	CBDV	<i>Babuvirus</i>	(99)
		SqLCV	<i>Begomovirus</i> (bipartite)	(64)	FBNYV	<i>Nanovirus</i>	(96)
		BCTV	<i>Curtovirus</i>	(75, 131)	MDV	<i>Nanovirus</i>	(96)
		TYLCV	<i>Begomovirus</i> (monopartite)	(132, 133)	SCSV	<i>Nanovirus</i>	(96)
		WDV	<i>Mastrevirus</i>	(71, 72, 136)	BBTV	<i>Babuvirus</i>	(97)
		ToLCV-Nde	<i>Begomovirus</i> (bipartite)	(76, 137)			
		MSV	<i>Mastrevirus</i>	(73)			
		MYMIV	<i>Begomovirus</i> (bipartite)	(70)			
	Rep oligomerisation	TGMV	<i>Begomovirus</i> (bipartite)	(66, 78)			
	The stem-loop sequence is not involved in recognition specificity of the viral DNA by the corresponding Rep protein	TGMV	<i>Begomovirus</i> (bipartite)	(74)			
		BGMV	<i>Begomovirus</i> (bipartite)	(74)			
	Replication initiation	Conserved stem-loop sequence is required for replication	TGMV	<i>Begomovirus</i> (bipartite)	(74, 82, 138)	BBTV	<i>Babuvirus</i>
SqLCV			<i>Begomovirus</i> (bipartite)	(64)			
MSV			<i>Mastrevirus</i>	(139)			
MDV			<i>Mastrevirus</i>	(140)			
BGMV			<i>Begomovirus</i> (bipartite)	(74)			
TYLCV			<i>Begomovirus</i> (monopartite)	(81)			

mechanisms of geminiviruses and nanovirids. We propose schematic representations for the different replication modes they may use. Importantly, we also highlight the fragmentary nature of our knowledge on this issue: i) there is no single species where all steps of the replication cycle have been studied and the complete picture stems from inferences across species (Tab. 1), ii) the quantitative importance of RCR versus RDR is unknown and the fate of replicative-intermediates is unclear, iii) the differential accumulation of genomic segments in multipartite species is not fully explained by the currently known mechanisms.

Rolling Circle Replication

1) Geminiviruses

Geminiviruses replication takes place in the nucleus of host cells (Nagar et al 1995), where they boost and divert the host DNA synthesis machineries. Some geminiviruses are restricted to phloem vascular tissues, while others are also present in mesophyll cells (25–27). All these cell types are fully differentiated and have therefore left the replicative stage of the cell cycle, implying that some factors required for viral DNA synthesis/replication are no longer produced. Geminiviruses can at least partially reprogram the cell cycle of their host in order to re-induce the replicative phase of the cell cycle and benefit from cellular actors necessary for DNA replication (28–30). This partial reprogramming is mediated by the interaction of geminiviral proteins with components of the cell cycle regulation. Interactions between the geminiviral proteins RepA or Rep and the host cell cycle regulator pRBR have been the most extensively studied. The RepA protein of mastreviruses, which is a replication enhancer, interacts with pRBR via a conserved LxCxE motif (31–36). In begomoviruses that do not encode a RepA protein, it is the Rep protein that can directly interact with pRBR (37). The Rep protein of begomoviruses lacks the LxCxE motif, and the interaction with pRBR takes place in the middle of an alpha helical motif located in the N-terminal region (38, 39). Geminivirus Rep proteins have also been found to interact with other host proteins involved in cell cycle regulation, as for example with the cellular proliferating cell nuclear antigen (PCNA) (40, 41), SUMO conjugation enzymes (41–43), a protein kinase and the histone H3 (44). Interactions between other geminivirus proteins and host proteins have been identified, as for example the geminiviral replication enhancer REn and PCNA (40), REn and pRBR (45) or C4 and a ubiquitin E3 ligase KRP (46), phosphorylation and SHAGGY-like kinases (47) and the NbSK η kinase (48).

		WDV	<i>Mastrevirus</i>	(136)			
	Rep binding leads to a distortion of the DNA structure at the origin of replication	MYMIV	<i>Begomovirus</i> (bipartite)	(70)			
	Rep cleavage in the nonanucleotide sequence	ACMV	<i>Begomovirus</i> (bipartite)	(58)	BBTV	<i>Babuvirus</i>	(100)
		TYLCV	<i>Begomovirus</i> (monopartite)	(81, 141)	FBNYV	<i>Nanovirus</i>	(95)
		WDV	<i>Mastrevirus</i>	(141, 142)			
New strand elongation	Rep protein remains covalently linked to the 5'-end of the cleaved positive single-strand	TYLCV	<i>Begomovirus</i> (monopartite)	(81)	BBTV	<i>Babuvirus</i>	(100)
	Rep helicase activity	TYLCSV	<i>Begomovirus</i> (monopartite)	(84)			
		MYMV	<i>Begomovirus</i> (bipartite)	(85)			
		MYMIV	<i>Begomovirus</i> (bipartite)	(85)			
		ICMV	<i>Begomovirus</i> (bipartite)	(85)			
New strand synthesis is performed by host polymerases δ or ϵ	TYLCV	<i>Begomovirus</i> (monopartite)	(59)				
Joining of the plus-strand extremities	Rep cleavage and joining activity	TYLCV	<i>Begomovirus</i> (monopartite)	(81, 141)	BBTV	<i>Babuvirus</i>	(100)
		WDV	<i>Mastrevirus</i>	(141)			

Shaded cells indicate that the corresponding step has not been investigated.

(i) Complementary-strand synthesis

The first step of replication is the synthesis of the complementary-strand (csDNA) to convert the viral ssDNA entering the host cell into a double-stranded DNA (dsDNA) on which replication will proceed. By studying incomplete intermediates of csDNA synthesis from leaves infected by African cassava mosaic virus (ACMV; *Begomovirus*, bipartite), Saunders and colleagues (49) determined that the csDNA synthesis is initiated into the CR, near a sequence that can potentially form a stem-loop structure. The csDNA intermediates are associated with RNA moieties, suggesting that csDNA synthesis may be preceded by the synthesis of an RNA primer (49).

Several studies consistently showed that small DNA fragments are associated with mastrevirus particles and are able to prime the csDNA synthesis *in vitro* (50–55). These DNA primer molecules are complementary to the Small Intergenic Region (SIR) of the viral-strand DNA. They are mainly about 80 nucleotides, but heterogeneous in length with a conserved 5'-end position and a variable 3'-end position. Their 5'-end is linked to a few ribonucleotides which are supposed to derive from a larger RNA.

These results, obtained partly on mastreviruses and partly on begomoviruses, together suggest that the csDNA synthesis is initiated into a conserved non-coding region (SIR for mastreviruses or CR for begomoviruses) *via* the synthesis of an RNA primer probably synthesized by a host primase. Recently, Wei and Lozano-Durán (56) investigated the role of the primase subunits PRIM1 and PRIM2 of the DNA polymerase α which is required for complementary-strand synthesis (see below). They determined that silencing of *PRIM1* or *PRIM2* genes in *Nicotiana benthamiana* virtually abolished the accumulation of the monopartite begomovirus tomato yellow leaf curl virus (TYLCV). However, the authors cautiously indicated that this effect could be due to a destabilization of the DNA polymerase α caused by the absence of one of its subunits, and that the direct involvement of primases thus awaited confirmation. In this direction, a quantitative trait locus (QTL) analysis determined that mutations in the *DNA Primase Large subunit (PRiL)* gene of *Cucumis melo* genome could confer resistance to the *Tomato leaf curl New Delhi virus* (ToLCNDV) (57). In the same study, the silencing of *N. benthamiana PriL* led to a severe reduction of accumulation of different geminiviruses, further supporting a role of primase subunits in the initiation of complementary-strand synthesis.

In some cases, and for unknown reasons, the csDNA synthesis could stop soon after its initiation (about 80 nucleotides), resulting in viral ssDNA molecules with a small double-stranded region as those found in mastreviruses virions. These partially dsDNA molecules could be encapsidated and move to a new cell or

host. The small complementary DNA fragments could then prime csDNA synthesis during a new infection event, after uncoating of the viral DNA (54). However, it is unclear whether this scheme applies to all geminiviruses since no DNA primers similar to those of mastreviruses were found associated with ACMV virions (58).

As the geminivirus genome does not encode a polymerase, the synthesis of the csDNA is performed by a cellular DNA polymerase complex. It has been demonstrated that polymerase α is essential for geminiviruses csDNA synthesis (59). Nevertheless, due to its low processivity, the authors did not exclude the involvement of another polymerase yet to be identified. Likewise, the possible involvement of translesion synthesis polymerases (TLS; *Pol* ζ and *Pol* η) has been mentioned (60). These enzymes are specialized DNA polymerases that can replicate damaged DNA templates by bypassing DNA lesions and have recently been shown to stabilize the genome of human dsDNA herpesvirus (61). The euphorbia yellow mosaic virus (EuYMV) could infect *Arabidopsis thaliana* lines knocked out for these genes, but deep sequencing revealed a significantly enhanced mutation rate. The authors concluded that TLS are not mandatory but that they may contribute to complementary strand synthesis, particularly at very early stages of infection (60).

Following the synthesis of the csDNA, the viral DNA is in a circular double-stranded form.

(ii) Recognition of the viral DNA by the Rep protein

The replication of the genome of geminiviruses is initiated by Rep proteins (Replication-associated proteins). They are not DNA polymerases, but a family of replication initiator proteins, promoting or controlling the RCR mechanism (62). Rep proteins bind the viral single- or double-stranded DNAs at the CR in a sequence-specific manner (63–66). The CR, which is variable among geminiviruses but identical between components DNA-A and DNA-B of bipartite begomoviruses, contains two key elements for replication: short iterative sequences called iterons and an inverted complementary sequence that can form a stem-loop structure.

Several studies have demonstrated that the specific recognition between viral genomic DNA and Rep protein is driven by iterons (Tab. 1). A comparative analysis of the intergenic regions from 30 geminiviruses (67) identified short iterative sequences of various sizes (Fig. S1A). The arrangement of these motifs is highly conserved among the different phylogenetic clades of geminiviruses, but their sequences are specific to viral species. The Rep protein of the tomato golden mosaic virus (TGMV) binds with high affinity to a GGTAGTAAGGTAG sequence, composed of two GGTAG motif repeats separated by TAA, located in the

CR upstream of the sequence potentially forming a stem-loop structure (68). The authors performed competition assays between the complete repeated sequence (GGTAGTAAGGTAG), the 5'-motif unit plus the spacer (GGTAGTAA) and the spacer plus the 3'-motif unit (TAAGGTAG). They determined that the Rep protein recognizes the combination of the two repeats. Rep binds preferentially to the 3'-motif but both repeats are necessary for high affinity. The Rep affinity is reduced by the replacement of the first two guanines by cytosines in the repeated motif and much more so when these mutations are engineered in the 3'-repeat (GGTAGTAAcCTAG) than the 5'-repeat (ccTAGTAAGGTAG). The same authors further investigated the importance of these iterons in viral replication by transfecting a fragment of the TGMV DNA-A sequence containing wild-type (GGTAGTAAGGTAG) or mutated (ccTAGTAAGGTAG; GGTAGTAAcCTAG) motifs with a Rep gene (*AL1*) expression vector in *Nicotiana tabacum* protoplasts. No newly synthesized DNA replicon accumulation was detected for the 3'-mutated iteron (GGTAGTAAcCTAG). Low level of viral accumulation was detected for the 5'-mutated iteron (ccTAGTAAGGTAG) and solely when a vector expressing the replication-activating gene *AL3* was added, but much lower than the wild-type sequence. Thus, the two iterons are involved in recognition by Rep but they do not have the same importance for viral replication. Additionally, the importance of the distance between the iterons and the stem-loop sequence was highlighted by a study on TGMV, where the addition of 7, 10 or 13 nucleotides between the two almost completely abolished viral replication, whereas the addition of only three nucleotides at the same location had little effect (69). The authors demonstrated that the deleterious effect on replication of the additional sequences was caused by the alteration of the iteron/stem-loop distance rather than by the nature of the inserted sequence. Though most studies located the iteron sequences of the members of the genus *Begomovirus* on the left side of the stem-loop sequence, it was recently demonstrated that the mung yellow mosaic India virus (MYMIV; *Begomovirus*, bipartite) genome has iteron sequence CGGTGTA repeated on both sides of the stem-loop sequence with two repeats on the left and one on the right (70). Moreover, studies conducted on two mastreviruses, the wheat dwarf virus (WDV; (71, 72)) and the maize streak virus (MSV; (73)) also showed Rep-binding sites on both sides of the stem-loop sequence.

Fontes and collaborators (74) analyzed the binding specificities of the two bipartite begomoviruses TGMV and bean golden mosaic virus (BGMV) Rep proteins *in vitro*. For this purpose, the authors swapped the iteron sequences of the TGMV and BGMV DNA-B components and analyzed the ability of the Rep protein of each virus to bind to the wild-type and mutated DNA-B. The TGMV Rep protein was able to bind to the mutated BGMV DNA-B component containing the iteron sequence of the TGMV, while the BGMV Rep protein did not. Conversely, the BGMV Rep protein bound to the mutated TGMV DNA-B containing

the BGMV iteron sequences, while the TGMV Rep protein did not. Despite efficient Rep fixation observed *in vitro*, no replication of the DNA-B mutants was observed when they were co-infiltrated with wild-type DNA-A of each virus in tobacco protoplasts. The authors concluded that Rep binding to iteron sequences is necessary for replication, but not sufficient. As the loop sequences of TGMV and BGMV differ by two nucleotides, the same authors then investigated the role of the stem-loop sequence in the recognition specificity of the viral DNAs by their cognate Rep protein. The authors mutated the BGMV DNA-B stem-loop sequence such that it corresponds to the TGMV DNA-B sequence. They then analyzed the accumulation of wild-type and mutated BGMV DNAs-B when co-transfected with BGMV or TGMV DNAs-A in tobacco protoplasts. They observed that the mutated BGMV DNA-B, which has the TGMV stem-loop sequence, is accumulated at a similar level to that of the wild-type BGMV DNA-B in presence of BGMV DNA-A. No replication was detected when transfected with TGMV DNA-A. The observed sequence variation in the stem loop sequences is thus not involved in recognition specificity of the viral DNA by the corresponding Rep protein.

Proteins of the rep family have three major conserved functional domains: endonuclease and DNA-binding, helicase, and oligomerization (1). While the helicase and endonuclease functions are evoked later in the step-by-step description of the replication process, the DNA binding motif is important at the very early phases. Specific recognition of the CR sequence by the protein Rep is controlled by the Rep N-terminal region (75, 76). A comparison of the iterons and the N-terminal region of the Rep protein of over a hundred geminiviruses identified an “iteron-related domain” (IRD) (77). This domain results from the assembly of Rep oligomers (66, 78) where a not yet delineated stretch of N-terminal amino acids is folded into a variable structure (79). Amino acid variations in this structure correlate with variations within the iteron sequences, suggesting that the IRD is an important Rep determinant of the specific recognition and replication initiation of a viral ssDNA circle (77, 80).

(iii) Replication initiation

The stem loop sequence being highly conserved in the non-coding region (Long Intergenic Region (LIR)/IR) of all geminiviruses, its role in replication has been further investigated (67, 74, 81). This sequence contains a central loop-forming nonanucleotide motif, TAATATTAC for most geminiviruses or TAAGATTCC for *Eragrovirus* and *Becurtovirus* (7), flanked by reverse complementary sequences constituting the stem that can vary according to the viral species (74, 81). Located into the highly conserved CR, the stem-loop sequence of bipartite begomoviruses is identical between DNA-A and -B. The Rep protein has a DNA cleavage activity between nucleotides T and A in bold of the nonanucleotide sequence TATTATT**AC** of the

TYLCV (*Begomovirus*, monopartite) (81). Nucleotide substitutions in this sequence (including changes of the nucleotides T and A in bold) did not affect the location of Rep cleavage, which remained between nucleotides seven and eight of the loop-sequence, but decreased its efficiency (*i.e.* the cleaved DNA quantity) (81). Contradictory to the fact that the viral DNA is double-stranded at this early stage of the replication cycle, the TYLCV Rep protein cleaves only single-stranded DNA and preferentially the positive one. The single-stranded negative strand can also be cleaved between the nucleotides in bold TAATATT/ATA or TAATATTAT/A, but the products are detected only after a long exposure time, indicating the lower efficiency of the reaction (81). The same authors (81) reported an efficient cleavage of oligonucleotides lacking the flanking sequence on either side of the loop-forming nonanucleotide motif, and concluded that the stem structure is not required. However, Orozco and Hanley Bowdoin (82) showed that the stem-loop structure formation is essential for TGMV replication *in vivo*. Most likely, the important role of the stem, which forms in mirror (Fig. 2A step 3), is to force the separation of the two strands of the loop-forming nonanucleotide, and thus render the Rep cleavage site accessible as single-stranded. Consistently, Singh and colleagues (70) indicated that MYMIV Rep binding to the iterons, located on both sides of the stem-loop sequence, distorts the dsDNA at the origin of replication (ORI). It is noticeable that a similar observation was reported for pT181 plasmid, which replicates in a rolling circle manner and where the RepC protein binding enhances the formation of the stem-loop structure *in vitro* and probably *in vivo* (81, 83). Together, these results suggest that the fixation of the Rep protein on the geminiviruses dsDNA ORI sequence induces the formation of the stem-loop structure. Rep is then able to efficiently cleave the positive single-stranded nonanucleotide loop, leaving the negative csDNA mostly uncleaved.

The sequence of the stem appears to be important for its facilitating role in loop-forming. Indeed, Orozco and Hanley-Bowdoin (82) introduced paired-mutations in order to change the sequence of the stem without affecting the structure formation in TGMV DNA-B. These modifications impaired DNA-B replication when expressed with a competing wild-type origin provided by DNA-A *in vivo*. An explanation could be that, although they do not prevent the formation of the stem-loop structure, these mutations could affect the stability of base-pairing and thus make it less competitive than the wild-type sequence. Consequently, and although the precise series of events initiating replication remains partly unclear, it likely involves both the stem-loop structure and the nature of the sequence composing it.

(iv) Cleavage, elongation and joining

After viral DNA recognition and cleavage in the conserved nonanucleotide sequence, Laufs and colleagues (81) showed that the TYLCV Rep protein remains covalently linked to the 5'-extremity of the

cleaved positive single-strand. According to the replication model proposed by these authors, the released 3'-extremity is then available to prime the synthesis of the new plus-strand, while the 5'-end linked to the Rep is displaced along the viral DNA. It has been demonstrated recently that the host DNA polymerase δ is recruited by the C3 replication enhancer protein of different geminiviruses to synthesize the new viral DNA strand (59). In C3 null mutants, it is the host polymerase ϵ which allows the synthesis of the new viral ssDNA.

Once the full-length DNA is polymerized, the newly synthesized nonanucleotide sequence is recognized and cleaved by the Rep protein. Finally, the 3'- and 5'- ends of the new synthesized DNA are joined by the Rep, constituting a free circular single-stranded DNA (81).

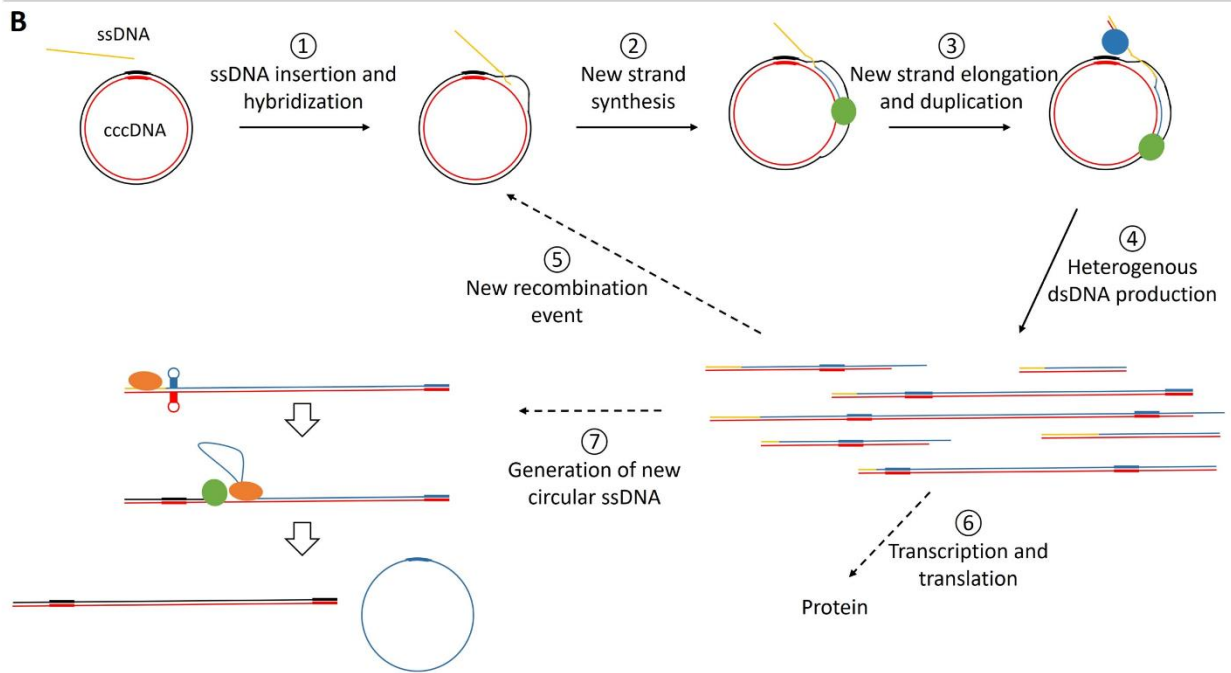
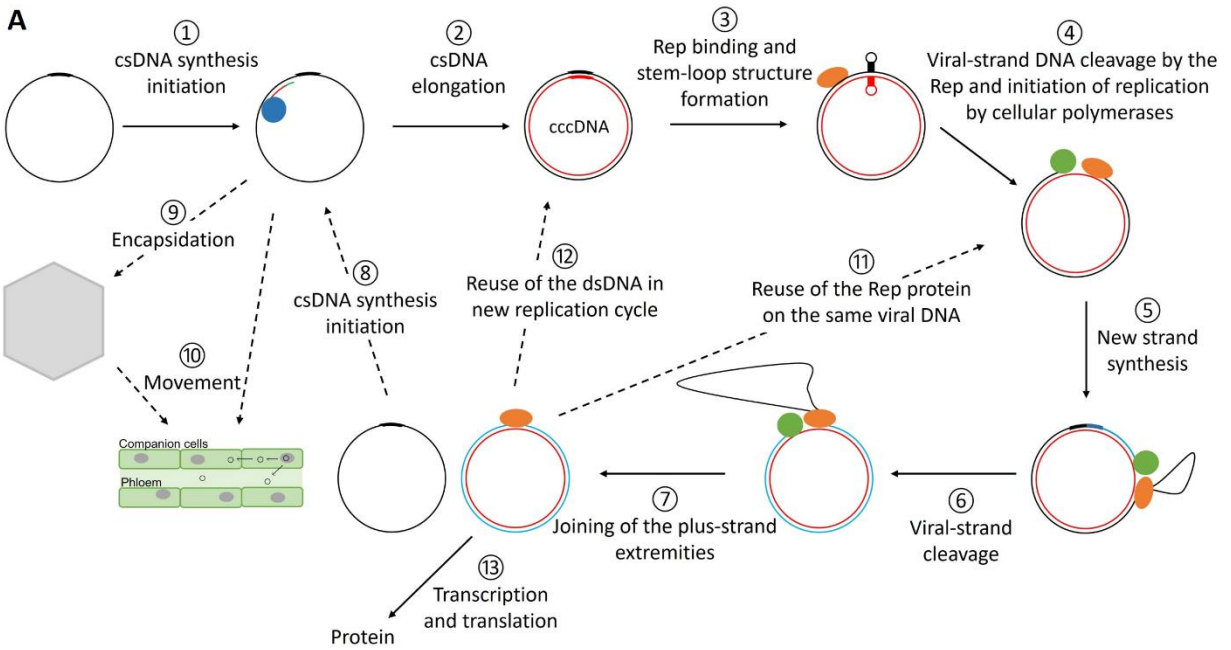
Although viral replication is ATP-dependent, the cleavage and ligation steps are not (81). The same study further suggested that the Rep protein may have helicase activity requiring ATP, like some Adeno-associated viruses. Consistent with this speculation, a helicase activity was later directly demonstrated *in vitro* for the Rep of the tomato yellow leaf curl Sardinia virus (TYLCSV, *Begomovirus*, monopartite) (84), and *in planta* for that of MYMIV (85). This latter study also highlighted a similar *in vitro* helicase activity for two additional begomoviruses, respectively close to (mung bean yellow mosaic virus) and more distant from (Indian cassava mosaic virus) MYMIV, suggesting that the helicase activity could be conserved among all geminivirus Rep proteins. Both studies showed that the geminivirus helicase activity is dependent on Rep proteins oligomerization and the unwinding is made from 3'- to 5'-end direction of the template strand.

(v) Integrated model for geminiviruses rolling circle replication

Based on the information provided by this literature, a diagram of geminiviruses replication *via* the RCR mechanism is proposed in Fig. 2. First, the csDNA synthesis is primed by an RNA or DNA primer in a conserved region (1). The csDNA elongation is then processed by the host cellular machinery, involving polymerase α (2). The Rep complex, composed of Rep proteins assembled as oligomers, binds the viral DNA at the iteron sequences that can be located either upstream or on both sides of the stem-loop sequence. The iteron sequences confer the specificity of recognition of the viral DNA by its Rep protein. The nucleotide sequence of the iterons itself, that of the flanking sequences, the spacing between the iterons and the spacing between iterons and stem-loop are all important. The binding of the Rep complex to the iteron sequences likely imposes physical constraints leading to a distortion of the DNA structure at the ORI, locally melting the double-strand and allowing the stem-loop structure formation (3). The

formation of this structure exposes a nonanucleotide single-stranded loop and the positive strand can then be preferentially cleaved by the Rep complex. After cleavage, Rep remains covalently linked to the 5'-end of the cleaved positive strand and the 3'-OH extremity serves as host polymerase primer (4). The synthesis of the new strand is performed by the host polymerases δ or ϵ , depending on whether the replication enhancer protein C3 is respectively present or not, using the csDNA synthesized in (2) as template (5). With its helicase activity, Rep unwinds the dsDNA from the 3'- to the 5'-end of the csDNA to allow the replication complex to progress. Once the new full-length DNA synthesized, Rep cuts the positive strand in the conserved nonanucleotide sequence, thus separating the replicated strand from the newly formed strand (6). Finally, Rep joins the two ends of the replicated positive strand to form a free circular DNA strand (7).

It is likely that a new csDNA initiation event occurs on the released circular ssDNA (8). The synthesis of the csDNA can be done entirely. In this case, the dsDNA can enter a new replication cycle (2). Otherwise, the elongation of csDNA could be stopped prematurely. The DNA molecules, with a small double-stranded portion, would be encapsidated (9) and/or move outside of the cell (10). The dsDNA generated by the RCR mechanism (7) could directly perpetuate the replication cycle by continuing to use the same Rep and polymerase complexes (true rolling circle replication, 11), or initiate a new cycle after recruiting new Rep and polymerase complexes (12). The dsDNA is also the form which is transcribed to produce viral RNAs and proteins (13). To our knowledge, the determinants of the distinct pathways and the proportion of the different RCR products following each pathway have not been characterized.



Stem-loop sequence
 Stem-loop structure
 Hypothesis
 primer
 Host polymerase δ or ϵ complex
 Host polymerase α complex
 Rep complex

Fig. 2: Replication modes of *Geminiviridae* and *Nanoviridae*

A- Schematic representation of rolling circle replication (RCR) of geminiviruses and nanoviruses. First, the complementary strand synthesis is initiated from a primer in a conserved region (SIR in mastreviruses, CR in begomoviruses; CR-M in babuviruses) of a viral single-stranded DNA (1). The elongation of the csDNA is processed by the host cellular machinery, notably polymerase α , and results in a circular covalently closed DNA molecule (cccDNA) (2). The Rep complex, constituted by Rep proteins assembled as oligomers, binds to the iteron sequences. The binding leads to a local melting of the dsDNA favoring the formation of a stem-loop structure (3). This stem-loop structure exposes a nanonucleotide single-stranded loop and permits the preferential cleavage of the positive strand by the Rep complex. Rep remains linked to the 5'-end of the cleaved strand and the 3'-OH extremity serves as primer for the host polymerase (4). Rep and the polymerase move through the viral DNA. With its helicase activity, Rep unwinds the dsDNA and the polymerase synthesizes the new positive strand using the complementary strand synthesized in (1) as a template (5). The positive strand is cut by the Rep in the conserved nonanucleotide sequence, thus separating the replicated strand (black strand) from the newly formed strand (blue strand) (6). The two extremities of the replicated positive strand are joined by the Rep to form a full genome length circular DNA (7). On the one hand, a new csDNA initiation event can occur on the released circular ssDNA (black; (8)). The csDNA can be completely synthesized and enter a new replication cycle (2). The elongation of csDNA could also be stopped prematurely and the generated DNA molecules could be encapsidated (9) and/or move outside of the cell (10). On the other hand, the dsDNA (7) could be used in a new replication cycle, either directly by recruiting a new polymerase complex (11), or by recruiting a new Rep protein (12). It could also be used to produce mRNAs and viral proteins (13).

B- Schematic representation of recombination-dependent replication (RDR) of geminiviruses. First, a viral ssDNA fragment (yellow) is inserted between the two strands of a cccDNA (1). The ssDNA fragment hybridizes with the homologous region inside the cccDNA and primes the synthesis of a new strand (blue) along the cccDNA template (2). During elongation, the new synthesized strand is converted into dsDNA by a host polymerase complex (3). The RDR mechanism leads to the formation of linear dsDNA molecules of various sizes (4). The complementary strand of some of them is not fully synthesized and their 3' overhang can initiate new RDR events (5). The linear dsDNA can be transcribed to produce viral proteins (6). Some of these linear dsDNAs contain several origins of replication and can serve as replication templates by Rep to produce a circular ssDNA in a way similar to RCR (7).

2) Nanovirids

Unlike geminiviruses, the replication of nanovirids has been little studied. Since their genome encodes a Rep protein and is constituted of circular genome segments each harboring non-coding regions similar to those described in geminiviruses, it is assumed that their replication also follows a rolling-circle mechanism. The few studies available explore different features of RCR in nanovirids in comparison with geminiviruses. While showing many similarities between the two virus families, they also point at some potential differences. As with geminiviruses, nanovirid replication takes place in the nucleus of the host cell. All nanovirids are restricted to phloem companion cells (86), where they likely also partially reprogram the cell into a replicative phase of the cell cycle. Consistently, the nanovirid Clink protein (for “cell cycle link”), possesses the LxCxE motif allowing its interaction with members of the pRB family, and hence stimulate viral replication (87–89). Clink also interacts with a S-phase kinase-associated protein 1 (KPN1) homologue (88). The nanoviral Rep proteins lack the LxCxE motif, and there are no reports of its direct involvement in host cell cycle manipulation.

(i) Complementary-strand synthesis

As for mastreviruses in the family *Geminiviridae*, DNA primers have been shown to be associated with BBTV (*Babuvirus*) segments extracted from purified virions (90). These primers have heterogeneous length, but the majority is about 80 nucleotides. All are initiated within the CR-M with a variable initiation site. Intriguingly, most of these primers extend beyond the end of the CR-M, into a non-conserved region where they show homology only to DNA-C. No ribonucleotides were found linked to the primers, but the authors of the study do not exclude a possible degradation of the RNA during the experimental procedure or removal by unknown processes before encapsidation. These primers are capable to prime csDNA synthesis of BBTV DNA components *in vitro*. Although the authors pointed out that the DNA obtained by virion purification contains similar amounts of each BBTV component, most of the self-primed products were derived from BBTV DNA-C, consistent with the sequences of the primers identified. Products derived from DNA-M, N, R and U3 were similarly weakly accumulated, but no products were detected for DNA-S. According to the authors, it is likely that DNA-S is capable of self-priming but that the reaction products are too rare to be detected. More recently, it has been shown that BBTV DNA components accumulate at different frequencies into banana plants and although DNA-C accumulates more than DNA-S, both are poorly accumulated compared to DNA-N, R and U3 (91, 92). This leaves the observation of an overrepresentation of segment C primers completely unexplained. DNA-C encodes the protein Clink, which interacts with cell cycle regulators to enhance replication (7, 88) and DNA-S encodes the coat protein

(93). Typically, during viral infection, genes encoding proteins involved in replication are first expressed. Those encoding structural proteins, such as coat proteins, are expressed later (94). Thus, there may be a link between complementary-strand synthesis (providing transcription template) and gene expression timing, where DNA-C would be preferentially expressed at early stages.

(ii) Recognition specificity

In nanovirids it is generally assumed that the recognition specificity of the virus DNA by its Rep protein is also determined by iterons. Iteron sequences are generally conserved between genomic segments of a given species. However, there must be some exceptions since an iteron located on the left side of the subterranean clover stunt virus (SCSV, *Nanovirus*) stem-loop sequence shows two different groups of sequences (DNA-C/M/U1/U4 and N/R/S/U2) (Fig. S1E). The nanovirids iterons have a different arrangement than those of geminiviruses. The sequences are not necessarily repeated in tandem, and some are inverted repeats. In addition, iteron sequences are always found on both sides of the stem-loop sequence, unlike for geminiviruses where iteron sequences are often found only upstream of the stem-loop (Fig. S1B-F; (95–99)). Introduction of mutations in some BBTV DNA-N iteron sequences affected its replication in banana embryogenic cells, and the quantitative effect depended on the iteron which was mutated (97). The ability of each of the faba bean necrotic yellows virus (FBNYV), milk vetch dwarf virus (MDV) and SCSV Rep proteins to initiate replication of DNA-S of heterologous species when agro-infiltrated in leaves of *Nicotiana benthamiana* has been studied (96). These three viruses share conserved sequences in the replication origin of their segments, especially FBNYV and MDV (Fig. S1B, D and E). Their Rep proteins also share similarities in their amino-acid sequences. By measuring segment accumulation in *N. benthamiana* leaves, the authors demonstrated that the Rep proteins of the three nanoviruses can drive the replication of heterologous nanoviruses DNA-S. However, the SCSV DNA-S accumulation is lower with the Rep of FBNYV and MDV than with its cognate Rep. Similarly, accumulation of MDV DNA-S with the SCSV Rep protein is lower than with the MDV or FBNYV Rep. This is perfectly in line with observed iteron sequences, for which FBNYV and MDV are more similar and SCSV more distant (Fig. S1B, D and E).

Due to both the poor characterization of iterons and the small number of available sequences, whether an IRD functional domain exists in nanovirus Rep proteins and specifically determines the ssDNA recognition and replication initiation is undocumented.

(iii) Replication of the positive strand

As in geminiviruses, the main conserved region of nanovirids contains a sequence potentially forming a stem-loop structure (Fig. 1B; (14–17)). The stem sequences vary between viral species, but also between genome segments of a species. For example, the faba bean necrotic stunt virus (FBNSV, *Nanovirus*) inverted-repeat sequences are identical for DNA-M/N/R/S/U2/U4 and DNA-C/U1 but differ between these two groups of segments. The loop also contains a nonanucleotide motif, conserved between the species of a given genus: AGTATT/AC in *Nanovirus* and ATTATT/AC in *Babuvirus* (95, 100); Fig. S1). As for geminiviruses, the nanoviruses Rep proteins possess a cleavage and joining activity. Rep cleaves the virion-sense DNA sequence in a single-stranded form, into the nonanucleotide motif between nucleotides T and A in bold (95, 100). After cleavage and during the elongation of the new strand, the Rep protein of BBTV remains covalently linked to the 5'-end of the positive single strand. Once the entire virus strand replicated, the Rep complex catalyzes the joining of the two extremities of the replicated strand (100).

No studies have yet investigated helicase activity in nanovirids. Nevertheless, the Rep amino acid sequence of FBNYV has a nucleoside triphosphate (NTP) binding motif (GxxGxxGKT/S), which suggests an ATPase activity essential for viral replication (95). By homology with geminiviruses, this ATPase activity could be required for a helicase function of nanovirids Rep proteins, but this has not yet been experimentally demonstrated.

(iv) Comparison of nanoviral and geminiviral RCR

In summary, nanovirids RCR starts by csDNA synthesis. As established for the mastreviruses of the family *Geminiviridae*, babuviruses have also been shown to encapsidate DNA primers enabling the priming of the csDNA synthesis. These primers match to a non-coding region, near a sequence that could potentially form a stem-loop structure. However, contrasting with geminiviruses and despite a similarity in size, the babuvirus primers have a variable 5'-end which does not seem to be linked to ribonucleotides. Perhaps biologically significant, but yet uncharacterized, these babuvirus primers extend into non-conserved regions conferring them a segment-specificity, which has not been reported in bipartite begomoviruses.

Once the csDNA synthesized, replication is initiated in a stem-loop sequence located in the conserved region CR-SL, which, as for geminiviruses, can differ in sequence between virus genera and species. A peculiarity of nanovirids is that the stem sequence can also vary between genome components of a species, while it appears to be the same for components DNA-A and DNA-B of bipartite geminiviruses. The

differences between stem-loop sequences of distinct segments of a given nanovirus species has thus far not been shown to have any functional significance, a point further discussed later.

In both families, the specificity of recognition of the viral DNA is driven by the iterons. Beyond their slightly different arrangement, the nanovirid iterons have always been identified on both sides of the stem-loop sequence while those of the geminiviruses are located either upstream or on both sides.

Finally, nanovirid Rep protein could possess an ATPase activity that could indicate a helicase function as shown in geminiviruses. Tab. 1 summarizes the similarities and differences between the RCR replication process of geminiviruses and nanovirids. Though some differences exist, RCR replication for geminiviruses and nanovirids is very similar. Therefore, the diagram presented in Fig. 2A could broadly apply to nanovirids as well.

Recombination-Dependent Replication

By studying the replicative intermediates of the bipartite begomovirus ACMV with two-dimensional agarose gel electrophoresis, Saunders and colleagues observed concatemeric linear and totally or partially double-stranded DNA forms not consistent with the RCR mechanism, for which open circular double-stranded DNAs, covalently closed double-stranded DNAs (cccDNA) and linear or circular single stranded DNAs are expected (20). Later, combining two-dimensional gel electrophoresis and electron microscopy, the group of Holger Jeske (101) determined that only a minority of abutilon mosaic virus (AbMV; *Begomovirus*, bipartite) DNA intermediates are compatible with RCR. Most of the DNA forms that were observed were linear dsDNAs without super-helicity, heterogeneous in length and in 5'- and 3'-ends on the AbMV genome. The authors concluded that these replication intermediates are consistent with the Recombination-Dependent Replication (RDR) mechanism, which is also found in some phages such as the dsDNA bacteriophage T4 (102). Nevertheless, as the authors pointed out, even if they suggested an RDR for AbMV, they did not completely prove it. After this seminal discovery, the DNA forms consistent with RDR were found in geminiviruses belonging to different genera, suggesting that this replication mechanism could be common in this virus family (103–107).

Assuming that this replication mode is indeed occurring, and based on phage replication knowledge, a model for RDR in geminiviruses is presented in Fig. 2B. Since it requires the involvement of cellular factors involved in DNA replication and in DNA damage repair, RDR must take place in the host cell nucleus. First,

a viral ssDNA fragment (yellow), probably originated from partially replicated ssDNA, is inserted anywhere between the two strands of a cccDNA generated by the synthesis of the csDNA (1). The ssDNA fragment hybridizes with the homologous region inside the cccDNA and primes the synthesis of a new strand (blue) along the cccDNA template (2). The elongation of the newly synthesized ssDNA continues along the circular double-stranded template during one or more rounds, possibly processed by host polymerase δ ; as shown for step 5 of RCR (Fig. 2A; (59)). During the new strand extension, the virus-strand (yellow then blue) is converted into dsDNA by a host polymerase complex (101, 108) (3). This step probably involves host polymerase α , since Wu and colleagues (59) determined that this enzyme is required to generate double-stranded intermediates in geminiviruses (Fig. 2A, step 1). RDR leads to the formation of linear dsDNA molecules of distinct sizes and different 5'- and 3'-ends according to the site of initiation and ending of elongation (4). The factor determining elongation termination is not known. Sometimes, the csDNA is not fully synthesized, and some extremities remain single-stranded. The 3'-overhang generated by these incomplete syntheses could be used to initiate new RDR events (5) (Pooggin, 2013). The linear dsDNA could be transcribed to produce viral proteins (6). Some of them, generated after several rounds around the circular DNA template, possess multiple origins of replication and could be used as replication templates by Rep to produce a circular ssDNA (108)) (7). Thus, the RDR mechanism does not use the ORI and does not involve the Rep protein, but for step 7. Interestingly, without this last step, the whole RDR process would not be a mode of replication but perhaps a way to produce transcription templates. More generally, in the absence of a better understanding of the processes controlling RDR it is difficult to appreciate the fate of its products and hence its role in the viral cycle functioning.

As earlier reported (109), geminivirus infection generates a genotoxic response activating the host DNA-repair machinery, promoting somatic homologous recombination (30, 110), which most likely induces the RDR mechanism. The TLS Pol ζ and Pol η , mentioned in a previous section for a possible role during complementary strand synthesis, may also contribute to RDR (Richter, Götz et al 2016). Most interestingly, a series of studies have highlighted the role of recombination mediator proteins in the replication of distinct geminiviruses. A phage display screen for *Arabidopsis thaliana* proteins binding to the mung bean yellow mosaic virus (MYMIV) Rep identified RAD54 (111) and RAD51 (112). In these studies, the interactions were confirmed using yeast two-hybrid and co-immunoprecipitation assays, and the requirement for replication was supported by *ex-vivo* approaches either in yeast or in infiltrated plant leaves monitoring geminivirus DNA replication in the presence/absence of RAD54 or RAD51. Two complementary studies further investigated the role of RAD54 (113) and several paralogues of RAD51 (114) in geminiviral infection. These *in vivo* studies used *Arabidopsis* gene knockouts challenged with

euphorbia yellow mosaic virus (EuYMV). While the involvement of RAD54 could not be confirmed *in vivo*, the paralogue RAD51D proved to promote viral replication at the early stages of infection. This activity was further shown to be linked to RAD51D earlier reported single-strand annealing (SSA) recombination capacity (115), indicating a putative important role in both the formation of replicative intermediates and the RDR process (114).

Viral intermediates from both RCR and presumed RDR are produced from the early stages and simultaneously during systemic infection of leaves with a geminivirus (101, 104, 116). These observations are not consistent with the replication mechanism of T4 phage for which the two mechanisms occur sequentially: RCR occurs in the early stages of infection, then, once viral accumulation reaches some threshold, the replication mode switches to RDR (104). The authors proposed that the apparent simultaneous production of intermediates of both replication modes could result from several asynchronous geminivirus cell infection events in the leaf tissues.

Thus far, no analogous study has been conducted for nanovirids, and whether they could also use RDR for viral DNA replication is totally unknown.

Concluding remarks

Numerous studies on geminivirus replication have allowed a good characterization of the RCR mechanism in these ssDNA viruses and permitted to establish a detailed diagram of RCR. Two genomic elements are crucial for RCR in *Geminiviridae*: (i) the iteron sequences that determine the specific recognition of the viral DNA by its Rep protein and (ii) the stem-loop sequence and structure that constitute the ORI. The few studies conducted in *Nanoviridae* suggest an RCR mechanism broadly similar to that of *Geminiviridae*. Furthermore, some intriguing observations indicate the possible existence of a second mode of replication by RDR. In contrast to RCR, this potential replication mode has been poorly studied in *Geminiviridae* and no studies have been conducted in *Nanoviridae*, leaving this process poorly understood.

Related to the multipartite architecture of nanovirids, many aspects of their life cycle remain enigmatic (117, 118), and replication is one of them. In addition to having their genome divided into a large number of genome segments, each of them accumulates at different relative frequencies within host plants. The frequency pattern of the distinct segments accumulating within infected plants, designated the genome

formula, is specific to the host species (119). It was recently showed (120) that the host related changes of FBNSV genome formulas are not induced by sequence modification. The same study highlighted a link between FBNSV genome formula and gene expression, suggesting that genome formula variations could allow FBNSV to adjust the expression of its genes upon host-switch by modulating their copy number. To date, the mechanisms of establishment and regulation of the genome formula are unknown. An obvious question is whether a differential replication of the distinct genome segments could explain their differential accumulation within the host plant. First, it is unlikely that differential segment accumulation results from different affinities with the replication complex because the iteron sequences allowing the recognition of the viral DNA by the Rep complex are highly conserved between different segments of the same nanovirus species. Second, the sequences of the stem allowing the formation of the stem-loop structure, hence permitting viral strand cleavage by the Rep complex, differ among segments. These differences could potentially affect replication efficiency by varying the distance between the Rep-binding site (iterons) and the cleavage site (loop) or by affecting the stem-loop structure stability. However, this cannot provide a satisfactory explanation either, because in the two groups of FBNSV segments M/N/R/S/U4 and C/U1, each containing segments with the exact same stem loop sequence (Fig. S1C; (121)), some are highly accumulated (*e.g.* N or U4 in *Vicia faba*) and others are rare (*e.g.* R and S in *V. faba*; (119, 120)). Third, RDR could occur preferentially on certain segments, which would generate differential accumulations. Again, this does not appear as a satisfactory explanation, as it could hardly explain the genome formula changes upon host-switching. Finally, initiation of the csDNA synthesis is a crucial step for replication since it allows the production of template for RCR and RDR. Interestingly, from the study by Hafner and colleagues (90), it appears that the synthesis of the BBTV csDNA is not equally primed over the six components *in vitro*, and this could be another explanation for unequal accumulation of the segments. In this study, most self-priming extension products corresponded to DNA-C, whereas no DNA-S extension could be detected. Different efficiency of primer association and encaspidation with distinct ssDNA segments could result in different production of csDNA early in infection and, consequently, in later accumulation building up the genome formula. This possibility has not been tested experimentally, and a parallel quantification of the efficiency of self-priming and of the genome formula in the same nanovirids species may provide an answer.

While the importance of RDR relative to RCR in the production of viral progeny remains unclear, RDR provides an interesting potential mechanism for the regulation of gene expression that has not been envisaged thus far. The large amounts of linear dsDNA produced by RDR are heterogeneous in length. They can be smaller or larger than the full-length genome, the larger ones appearing as a size continuum rather

than as various size-classes each corresponding to a round number of full-length genome units. The accumulation of these heterogeneous dsDNAs should thus lead to an imbalance in the copy number of the different viral genes, as they are not equally present on these molecules. That the copy number of a gene directly impacts its expression level is extensively documented in all organisms, including nanoviruses ((120) and references within). Therefore, both the probable distinct relative frequency of the different genes in this RDR-generated population of geminiviral dsDNA, and the fact that they can be transcribed, most likely impacts on gene expression patterns. One could then speculate on a functional role of RDR, other than replication. By analogy to the functional role of the genome formula of the nanovirus FBNSV (120), RDR could be a way to implement Amplification-Mediated Gene Expression Tuning (AMGET; (122)) in geminiviruses, in both monopartite and bipartite species.

This hypothesis, however, should be considered with further nuances. First, transcriptional regulation of gene expression by viral and host proteins has been reported in several instances, imposing that the gene copy number variations could only be one additional regulation level. For example, TGMV Rep protein transcriptionally represses its own gene (123); the transcription of the downstream TrAP and REn genes depends on the suppression of Rep transcription (124); and the CP gene expression may be regulated in several geminivirus species through a conserved transcription factor binding site (124–127). It is here relevant to note that, in such a situation of a transcriptional regulatory network connecting several genes, even minute copy number variations are predicted to have dramatic non-linear positive or negative effects on gene expression, potentially inducing bifurcation in the behavior of the transcriptional regulatory network (128). Second, because the linear dsDNA products of RDR are methylated (129), the genes they contain might not be Pol-II-transcribed into mRNA. However, possible transcription of these RDR products by RNA pol-II has been envisaged (108). Pol-II transcription depends on the number and location of methylated sites, and this issue has not been addressed yet specifically for viral linear dsDNA (129, 130), leaving the question open for future investigation.

Acknowledgments

This work was funded by the French ANR : Nanovirus ANR-18-CE92-0028, as well as by the University of Montpellier MUSE, project MULTIVIR. MB acknowledges support from University of Montpellier, IRD, CNRS and INRAE. SB acknowledges support from INRAE and Department SPE. YM acknowledges support from CNRS and IRD. We thank two anonymous reviewers for their useful comments.

References

1. Tarasova E, Khayat R. 2021. A Structural Perspective of Reprs from CRESS-DNA Viruses and Their Bacterial Plasmid Homologues. *Viruses* 14:37.
2. Kazlauskas D, Varsani A, Koonin EV, Krupovic M. 2019. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nat Commun* 10:3425.
3. Rojas MR, Macedo MA, Maliano MR, Soto-Aguilar M, Souza JO, Briddon RW, Kenyon L, Rivera Bustamante RF, Zerbini FM, Adkins S, Legg JP, Kvarnheden A, Wintermantel WM, Sudarshana MR, Peterschmitt M, Lapidot M, Martin DP, Moriones E, Inoue-Nagata AK, Gilbertson RL. 2018. World Management of Geminiviruses. *Annu Rev Phytopathol* 56:637–677.
4. Varma A, Malathi VG. 2003. Emerging geminivirus problems: A serious threat to crop production. *Ann Appl Biol* 142:145–164.
5. Rwahnih MA, Rowhani A, Golino DA, Islas CM, Preece JE, Sudarshana MR. 2015. Detection and genetic diversity of Grapevine red blotch-associated virus isolates in table grape accessions in the National Clonal Germplasm Repository in California. *Can J Plant Pathol* 37:130–135.
6. Perry KL, McLane H, Thompson JR, Fuchs M. 2018. A novel grablovirus from non-cultivated grapevine (*Vitis* sp.) in North America. *Arch Virol* 163:259–262.
7. Fiallo-Olivé E, Lett J-M, Martin DP, Roumagnac P, Varsani A, Zerbini FM, Navas-Castillo J. 2021. ICTV Virus Taxonomy Profile: Geminiviridae 2021: This article is part of the ICTV Virus Taxonomy Profiles collection. *J Gen Virol* 102.
8. Hesketh EL, Saunders K, Fisher C, Potze J, Stanley J, Lomonosoff GP, Ranson NA. 2018. The 3.3 Å structure of a plant geminivirus using cryo-EM. *Nat Commun* 9:2369.
9. Saunders K, Richardson J, Lawson DM, Lomonosoff GP. 2020. Requirements for the Packaging of Geminivirus Circular Single-Stranded DNA: Effect of DNA Length and Coat Protein Sequence. *Viruses* 12:1235.
10. Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D. 1999. Geminiviruses: Models for Plant DNA Replication, Transcription, and Cell Cycle Regulation. *Crit Rev Plant Sci* 18:71–106.

11. Lazarowitz SG, Shepherd RJ. 1992. Geminiviruses: Genome structure and gene function. *Crit Rev Plant Sci* 11:327–349.
12. Lal A, Vo TTB, Sanjaya I,GNPW, Ho PT, Kim J-K, Kil E-J, Lee S. 2020. Nanovirus Disease Complexes: An Emerging Threat in the Modern Era. *Front Plant Sci* 11:558403.
13. Qazi J. 2016. Banana bunchy top virus and the bunchy top disease. *J Gen Plant Pathol* 82:2–11.
14. Boevink P, Chu PW, Keese P. 1995. Sequence of subterranean clover stunt virus DNA: affinities with the geminiviruses. *Virology* 207:354–361.
15. 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.
16. Katul L, Maiss E, Morozov SY, Vetten HJ. 1997. Analysis of six DNA components of the faba bean necrotic yellows virus genome and their structural affinity to related plant virus genomes. *Virology* 233:247–259.
17. Sano Y, Wada M, Hashimoto Y, Matsumoto T, Kojima M. 1998. Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome. *J Gen Virol* 79:3111–3118.
18. Beetham PR, Hafner GJ, Harding RM, Dale JL. 1997. Two mRNAs are transcribed from banana bunchy top virus DNA-1. *J Gen Virol* 78:229–236.
19. Thomas JE, Gronenborn B, Harding RM, Mandal B, Grigoras I, Randles JW, Sano Y, Timchenko T, Vetten HJ, Yeh H-H, Ziebell H, ICTV Report Consortium. 2021. ICTV Virus Taxonomy Profile: Nanoviridae. *J Gen Virol* 102.
20. Saunders K, Lucy A, Stanley J. 1991. DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Res* 19:2325–2330.
21. Stenger DC, Revington GN, Stevenson MC, Bisaro DM. 1991. Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. *Proc Natl Acad Sci* 88:8029–8033.
22. Rosario K, Duffy S, Breitbart M. 2012. A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics. *Arch Virol* 157:1851–1871.
23. Ruiz-Masó JA, Machón C, Bordanaba-Ruiseco L, Espinosa M, Coll M, Del Solar G. 2015. Plasmid Rolling-Circle Replication. *Microbiol Spectr* 3:3.1.16.

24. Weigel C, Seitz H. 2006. Bacteriophage replication modules. *FEMS Microbiol Rev* 30:321–381.
25. Horns T, Jeske H. 1991. Localization of Abutilon Mosaic Virus (AbMV) DNA within leaf tissue by in situ hybridization. *Virology* 181:580–588.
26. Lucy AP, Boulton M, Davies JW, Maule A. 1996. Tissue Specificity of *Zea mays* Infection by Maize Streak Virus. *Mol Plant Microbe Interact* 9:22.
27. Wege C, Saunders K, Stanley J, Jeske H. 2001. Comparative Analysis of Tissue Tropism of Bipartite Geminiviruses. *J Phytopathol* 149:359–368.
28. Nagar S, Pedersen TJ, Carrick KM, Hanley-Bowdoin L, Robertson D. 1995. A geminivirus induces expression of a host DNA synthesis protein in terminally differentiated plant cells. *Plant Cell* 7:705–719.
29. Egelkrout EM, Robertson D, Hanley-Bowdoin L. 2001. Proliferating Cell Nuclear Antigen Transcription Is Repressed through an E2F Consensus Element and Activated by Geminivirus Infection in Mature Leaves. *PLANT CELL ONLINE* 13:1437–1452.
30. Ascencio-Ibáñez JT, Sozzani R, Lee T-J, Chu T-M, Wolfinger RD, Cella R, Hanley-Bowdoin L. 2008. Global Analysis of Arabidopsis Gene Expression Uncovers a Complex Array of Changes Impacting Pathogen Response and Cell Cycle during Geminivirus Infection. *Plant Physiol* 148:436–454.
31. Xie Q, Suárez-López P, Gutiérrez C. 1995. Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: requirement for efficient viral DNA replication. *EMBO J* 14:4073–4082.
32. Grafi G, Burnett RJ, Helentjaris T, Larkins BA, DeCaprio JA, Sellers WR, Kaelin WG. 1996. A maize cDNA encoding a member of the retinoblastoma protein family: involvement in endoreduplication. *Proc Natl Acad Sci* 93:8962–8967.
33. Xie Q, Sanz-Burgos AP, Hannon GJ, Gutiérrez C. 1996. Plant cells contain a novel member of the retinoblastoma family of growth regulatory proteins. *EMBO J* 15:4900–4908.
34. Horváth GV, Pettkó-Szandtner A, Nikovics K, Bilgin M, Boulton M, Davies JW, Gutiérrez C, Dudits D. 1998. Prediction of functional regions of the maize streak virus replication-associated proteins by protein-protein interaction analysis. *Plant Mol Biol* 38:699–712.

35. Liu L, Saunders K, Thomas CL, Davies JW, Stanley J. 1999. Bean Yellow Dwarf Virus RepA, but Not Rep, Binds to Maize Retinoblastoma Protein, and the Virus Tolerates Mutations in the Consensus Binding Motif. *Virology* 256:270–279.
36. Gutierrez C, Ramirez-Parra E, Mar Castellano M, Sanz-Burgos AP, Luque A, Missich R. 2004. Geminivirus DNA replication and cell cycle interactions. *Vet Microbiol* 98:111–119.
37. Ach RA, Durfee T, Miller AB, Taranto P, Hanley-Bowdoin L, Zambryski PC, Grissem W. 1997. *RRB1* and *RRB2* Encode Maize Retinoblastoma-Related Proteins That Interact with a Plant D-Type Cyclin and Geminivirus Replication Protein. *Mol Cell Biol* 17:5077–5086.
38. Kong L-J, Orozco BM, Roe J, Nagar S, Ou S, Feiler H, Durfee T, Miller AB, Grissem W, Robertson D, Hanley-Bowdoin L. 2000. A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *EMBO J* 19:3485–3495.
39. Arguello-Astorga G, Lopez-Ochoa L, Kong L-J, Orozco BM, Settlege SB, Hanley-Bowdoin L. 2004. A Novel Motif in Geminivirus Replication Proteins Interacts with the Plant Retinoblastoma-Related Protein. *J Virol* 78:4817–4826.
40. Castillo AG, Collinet D, Deret S, Kashoggi A, Bejarano ER. 2003. Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). *Virology* 312:381–394.
41. Arroyo-Mateos M, Sabarit B, Maio F, Sánchez-Durán MA, Rosas-Díaz T, Prins M, Ruiz-Albert J, Luna AP, Van Den Burg HA, Bejarano ER. 2018. Geminivirus Replication Protein Impairs SUMO Conjugation of Proliferating Cellular Nuclear Antigen at Two Acceptor Sites. *J Virol* 92:e00611-18.
42. Castillo AG, Kong LJ, Hanley-Bowdoin L, Bejarano ER. 2004. Interaction between a Geminivirus Replication Protein and the Plant Sumoylation System. *J Virol* 78:2758–2769.
43. Sánchez-Durán MA, Dallas MB, Ascencio-Ibañez JT, Reyes MI, Arroyo-Mateos M, Ruiz-Albert J, Hanley-Bowdoin L, Bejarano ER. 2011. Interaction between Geminivirus Replication Protein and the SUMO-Conjugating Enzyme Is Required for Viral Infection. *J Virol* 85:9789–9800.

44. Kong L-J, Hanley-Bowdoin L. 2002. A Geminivirus Replication Protein Interacts with a Protein Kinase and a Motor Protein That Display Different Expression Patterns during Plant Development and Infection. *Plant Cell* 14:1817–1832.
45. Settlege SB, Miller AB, Gruissem W, Hanley-Bowdoin L. 2001. Dual Interaction of a Geminivirus Replication Accessory Factor with a Viral Replication Protein and a Plant Cell Cycle Regulator. *Virology* 279:570–576.
46. Lai J, Chen H, Teng K, Zhao Q, Zhang Z, Li Y, Liang L, Xia R, Wu Y, Guo H, Xie Q. 2009. RKP, a RING finger E3 ligase induced by BSCTV C4 protein, affects geminivirus infection by regulation of the plant cell cycle. *Plant J* 57:905–917.
47. Mills-Lujan K, Andrews DL, Chou C, Deom CM. 2015. The Roles of Phosphorylation and SHAGGY-Like Protein Kinases in Geminivirus C4 Protein Induced Hyperplasia. *PLOS ONE* 10:e0122356.
48. Mei Y, Yang X, Huang C, Zhang X, Zhou X. 2018. Tomato leaf curl Yunnan virus-encoded C4 induces cell division through enhancing stability of Cyclin D 1.1 via impairing NbSK η -mediated phosphorylation in *Nicotiana benthamiana*. *PLOS Pathog* 14:e1006789.
49. Saunders K, Lucy A, Stanley J. 1992. RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. *Nucleic Acids Res* 20:6311–6315.
50. Donson J, Morris-Krsinich BA, Mullineaux PM, Boulton MI, Davies JW. 1984. A putative primer for second-strand DNA synthesis of maize streak virus is virion-associated. *EMBO J* 3:3069–3073.
51. Donson J, Accotto GP, Boulton MI, Mullineaux PM, Davies JW. 1987. The nucleotide sequence of a geminivirus from *Digitaria sanguinalis*. *Virology* 161:160–169.
52. Howell SH. 1984. Physical structure and genetic organisation of the genome of maize streak virus (Kenyan isolate). *Nucleic Acids Res* 12:7359–7375.
53. Andersen MT, Richardson KA, Harbison S-A, Morris BAM. 1988. Nucleotide sequence of the geminivirus chloris striate mosaic virus. *Virology* 164:443–449.
54. Hayes RJ, Macdonald H, Coutts RHA, Buck KW. 1988. Priming of Complementary DNA Synthesis in Vitro by Small DNA Molecules Tightly Bound to Virion DNA of Wheat Dwarf Virus. *J Gen Virol* 69:1345–1350.

55. Morris BAM, Richardson KA, Haley A, Zhan X, Thomas JE. 1992. The nucleotide sequence of the infectious cloned dna component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants. *Virology* 187:633–642.
56. Wei H, Lozano-Durán R. 2023. The primase subunits of DNA polymerase α , PRIM1 and PRIM2, are required for the replication of the geminivirus tomato yellow leaf curl virus in the host plant.
57. Siskos L, Antoniou M, Riado J, Enciso M, Garcia C, Liberti D, Esselink D, Baranovskiy AG, Tahirov TH, Visser RGF, Kormelink R, Bai Y, Schouten HJ. 2023. DNA primase large subunit is an essential plant gene for geminiviruses, putatively priming viral ss-DNA replication. *Front Plant Sci* 14:1130723.
58. Stanley J, Townsend R. 1985. Characterisation of DNA forms associated with cassava latent virus infection. *Nucleic Acids Res* 13:2189–2206.
59. Wu M, Wei H, Tan H, Pan S, Liu Q, Bejarano ER, Lozano-Durán R. 2021. Plant DNA polymerases α and δ mediate replication of geminiviruses. *Nat Commun* 12:2780.
60. Richter KS, Götz M, Winter S, Jeske H. 2016. The contribution of translesion synthesis polymerases on geminiviral replication. *Virology* 488:137–148.
61. Zeltzer S, Longmire P, Svoboda M, Bosco G, Goodrum F. 2022. Host translesion polymerases are required for viral genome integrity. *Proc Natl Acad Sci* 119:e2203203119.
62. Kazlauskas D, Varsani A, Krupovic M. 2018. Pervasive Chimerism in the Replication-Associated Proteins of Uncultured Single-Stranded DNA Viruses. *Viruses* 10:187.
63. Fontes EPB, Luckow VA, Hanley-Bowdoin L. 1992. A Geminivirus Replication Protein Is a Sequence-Specific DNA Binding Protein. *Plant Cell* 4:597.
64. Lazarowitz SG, Wu LC, Rogers SG, Elmer JS. 1992. Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. *Plant Cell* 4:799–809.
65. Thömmes P, Osman T a. M, Hayes RJ, Buck KW. 1993. TGMV replication protein AL1 preferentially binds to single-stranded DNA from the common region. *FEBS Lett* 319:95–99.

66. Orozco BM, Kong L-J, Batts LA, Elledge S, Hanley-Bowdoin L. 2000. The multifunctional character of a geminivirus replication protein is reflected by its complex oligomerization properties. *J Biol Chem* 275:6114–6122.
67. 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.
68. Fontes EP, Eagle PA, Sipe PS, Luckow VA, Hanley-Bowdoin L. 1994. Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *J Biol Chem* 269:8459–8465.
69. Orozco BM, Gladfelter HJ, Settlege SB, Eagle PA, Gentry RN, Hanley-Bowdoin L. 1998. Multiple cis elements contribute to geminivirus origin function. *Virology* 242:346–356.
70. Singh DK, Malik PS, Choudhury NR, Mukherjee SK. 2008. MYMIV replication initiator protein (Rep): Roles at the initiation and elongation steps of MYMIV DNA replication. *Virology* 380:75–83.
71. Castellano MM, Sanz-Burgos AP, Gutiérrez C. 1999. Initiation of DNA replication in a eukaryotic rolling-circle replicon: identification of multiple DNA-protein complexes at the geminivirus origin 1 1Edited by I. B. Holland. *J Mol Biol* 290:639–652.
72. Missich R, Ramirez-Parra E, Gutierrez C. 2000. Relationship of Oligomerization to DNA Binding of Wheat Dwarf Virus RepA and Rep Proteins. *Virology* 273:178–188.
73. Willment JA, Martin DP, Palmer KE, Schnippenkoetter WH, Shepherd DN, Rybicki EP. 2007. Identification of long intergenic region sequences involved in maize streak virus replication. *J Gen Virol* 88:1831–1841.
74. Fontes EP, Gladfelter HJ, Schaffer RL, Petty IT, Hanley-Bowdoin L. 1994. Geminivirus replication origins have a modular organization. *Plant Cell* 6:405–416.
75. Choi I-R, Stenger DC. 1995. Strain-Specific Determinants of Beet Curly Top Geminivirus DNA Replication. *Virology* 206:904–912.
76. Chatterji A, Padidam M, Beachy RN, Fauquet CM. 1999. Identification of Replication Specificity Determinants in Two Strains of Tomato Leaf Curl Virus from New Delhi. *J Virol* 73:5481–5489.

77. Argüello-Astorga GR, Ruiz-Medrano R. 2001. An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol* 146:1465–1485.
78. Orozco BM, Miller AB, Settlege SB, Hanley-Bowdoin L. 1997. Functional domains of a geminivirus replication protein. *J Biol Chem* 272:9840–9846.
79. Avalos-Calleros JA, Pastor-Palacios G, Bolaños-Martínez OC, Mauricio-Castillo A, Gregorio-Jorge J, Martínez-Marrero N, Bañuelos-Hernández B, Méndez-Lozano J, Arguello-Astorga GR. 2021. Two strains of a novel begomovirus encoding Rep proteins with identical β 1 strands but different β 5 strands are not compatible in replication. *Arch Virol* 166:1691–1709.
80. Gregorio-Jorge J, Bernal-Alcocer A, Bañuelos-Hernández B, Alpuche-Solís ÁG, Hernández-Zepeda C, Moreno-Valenzuela O, Frías-Treviño G, Argüello-Astorga GR. 2010. Analysis of a new strain of Euphorbia mosaic virus with distinct replication specificity unveils a lineage of begomoviruses with short Rep sequences in the DNA-B intergenic region. *Virol J* 7:275.
81. 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. *Proc Natl Acad Sci* 92:3879–3883.
82. Orozco BM, Hanley-Bowdoin L. 1996. A DNA structure is required for geminivirus replication origin function. *J Virol* 70:148–158.
83. Noirot P, Bargonetti J, Novick RP. 1990. Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. *Proc Natl Acad Sci* 87:8560–8564.
84. Clérot D, Bernardi F. 2006. DNA helicase activity is associated with the replication initiator protein Rep of tomato yellow leaf curl geminivirus. *J Virol* 80:11322–11330.
85. Choudhury NR, Malik PS, Singh DK, Islam MN, Kaliappan K, Mukherjee SK. 2006. The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. *Nucleic Acids Res* 34:6362–6377.
86. Chu PW, Boevink P, Surin B, Larkin P, Keese P, Waterhouse PM. 1995. Non-germinated single-stranded DNA plant viruses. *Pathog Host Specif Plant Disease Vol III Viruses Viroids* R.P. Singh, U.S. Singh and K. Kohmoto (eds) (Oxford: Elsevier Science Ltd).

87. 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. *J Gen Virol* 79:3101–3109.
88. 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* 74:2967–2972.
89. 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. *J Virol* 81:4177–4185.
90. Hafner GJ, Harding RM, Dale JL. 1997. A DNA primer associated with banana bunchy top virus. *J Gen Virol* 78:479–486.
91. Bashir S, Farrakh S, Yasmin T, Muhammad A, Bashir T, Manghwar H, Mora-Poblete F, Iqbal S, Baazeem A, Hyder MZ. 2022. Quantitation of Multipartite Banana Bunchy Top Virus Genomic Components and Their Transcripts in Infected Tissues of Banana (*Musa acuminata*). *Agronomy* 12:2990.
92. Guyot V, Rajeswaran R, Chu HC, Karthikeyan C, Laboureau N, Galzi S, Mukwa LFT, Krupovic M, Kumar PL, Iskra-Caruana M-L, Pooggin MM. 2022. A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids. *PLOS Pathog* 18:e1010448.
93. Wanitchakorn R, Harding RM, Dale JL. 1997. Banana bunchy top virus DNA-3 encodes the viral coat protein. *Arch Virol* 142:1673–1680.
94. Shimada-Beltran H, Rivera-Bustamante RF. 2007. Early and late gene expression in pepper huasteco yellow vein virus. *J Gen Virol* 88:3145–3153.
95. Timchenko T, Kouchkovsky F de, 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* 73:10173–10182.
96. 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.

97. 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.
98. Timchenko T, Bernardi F. 2007. Les nanovirus, petits virus de plantes :similitudes et différences avec les géminivirus. *Virologie* 11:27–42.
99. Stainton D, Martin DP, Collings DA, Varsani A. 2017. Comparative analysis of common regions found in babuviruses and alphasatellite molecules. *Arch Virol* 162:849–855.
100. Hafner GJ, Stafford MR, Wolter LC, Harding RM, Dale JL. 1997. Nicking and joining activity of banana bunchy top virus replication protein in vitro. *J Gen Virol* 78:1795–1799.
101. Jeske H, Lütgemeier, Preiss. 2001. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO J* 20:6158–6167.
102. Kreuzer KN. 2000. Recombination-dependent DNA replication in phage T4. *Trends Biochem Sci* 25:165–173.
103. Pilartz M, Jeske H. 2003. Mapping of Abutilon Mosaic Geminivirus Minichromosomes. *J Virol* 77:10808–10818.
104. Preiss W, Jeske H. 2003. Multitasking in replication is common among geminiviruses. *J Virol* 77:2972–2980.
105. Alberter B, Ali Rezaian M, Jeske H. 2005. Replicative intermediates of Tomato leaf curl virus and its satellite DNAs. *Virology* 331:441–448.
106. Jovel J, Preiß W, Jeske H. 2007. Characterization of DNA intermediates of an arising geminivirus. *Virus Res* 130:63–70.
107. Erdmann JB, Shepherd DN, Martin DP, Varsani A, Rybicki EP, Jeske H. 2010. Replicative intermediates of maize streak virus found during leaf development. *J Gen Virol* 91:1077–1081.
108. Pooggin M. 2013. How Can Plant DNA Viruses Evade siRNA-Directed DNA Methylation and Silencing? *Int J Mol Sci* 14:15233–15259.
109. Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S. 2013. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat Rev Microbiol* 11:777–788.

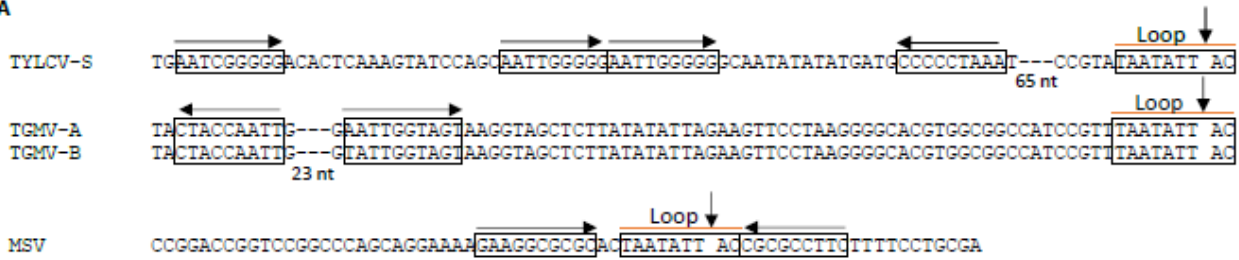
110. Richter KS, Kleinow T, Jeske H. 2014. Somatic homologous recombination in plants is promoted by a geminivirus in a tissue-selective manner. *Virology* 452–453:287–296.
111. Kaliappan K, Choudhury NR, Suyal G, Mukherjee SK. 2012. A novel role for RAD54: this host protein modulates geminiviral DNA replication. *FASEB J* 26:1142–1160.
112. Suyal G, Mukherjee SK, Choudhury NR. 2013. The host factor RAD51 is involved in mungbean yellow mosaic India virus (MYMIV) DNA replication. *Arch Virol* 158:1931–1941.
113. Richter KS, Ende L, Jeske H. 2015. Rad54 is not essential for any geminiviral replication mode in planta. *Plant Mol Biol* 87:193–202.
114. Richter KS, Serra H, White CI, Jeske H. 2016. The recombination mediator RAD51D promotes geminiviral infection. *Virology* 493:113–127.
115. Serra H, Da Ines O, Degroote F, Gallego ME, White CI. 2013. Roles of XRCC2, RAD51B and RAD51D in RAD51-Independent SSA Recombination. *PLoS Genet* 9:e1003971.
116. Paprotka T, Deuschle K, Pilartz M, Jeske H. 2015. Form follows function in geminiviral minichromosome architecture. *Virus Res* 196:44–55.
117. Sicard A, Michalakis Y, Gutiérrez S, Blanc S. 2016. The Strange Lifestyle of Multipartite Viruses. *PLOS Pathog* 12:e1005819.
118. Michalakis Y, Blanc S. 2020. The Curious Strategy of Multipartite Viruses. *Annu Rev Virol* 7:203–218.
119. 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. *Nat Commun* 4:2248.
120. Gallet R, Di Mattia J, Ravel S, Zeddani J-L, Vitalis R, Michalakis Y, Blanc S. 2022. Gene copy number variations at the within-host population level modulate gene expression in a multipartite virus. *Virus Evol* 8:veac058.
121. Grigoras I, Timchenko T, Katul L, Grande-Pérez A, Vetten H-J, Gronenborn B. 2009. Reconstitution of Authentic Nanovirus from Multiple Cloned DNAs. *J Virol* 83:10778–10787.

122. Tomanek I, Grah R, Lagator M, Andersson AMC, Bollback JP, Tkačik G, Guet CC. 2020. Gene amplification as a form of population-level gene expression regulation. *Nat Ecol Evol* 4:612–625.
123. Sunter G, Hartitz MD, Bisaro DM. 1993. Tomato Golden Mosaic Virus Leftward Gene Expression: Autoregulation of Geminivirus Replication Protein. *Virology* 195:275–280.
124. Shung C-Y, Sunter G. 2007. AL1-dependent repression of transcription enhances expression of Tomato golden mosaic virus AL2 and AL3. *Virology* 364:112–122.
125. Hofer JM, Dekker EL, Reynolds HV, Woolston CJ, Cox BS, Mullineaux PM. 1992. Coordinate regulation of replication and virion sense gene expression in wheat dwarf virus. *Plant Cell* 4:213–223.
126. Eagle PA, Orozco BM, Hanley-Bowdoin L. 1994. A DNA sequence required for geminivirus replication also mediates transcriptional regulation. *Plant Cell* 6:1157–1170.
127. Sardo L, Lucioli A, Tavazza M, Masenga V, Tavazza R, Accotto GP, Noris E. 2011. An RGG sequence in the replication-associated protein (Rep) of Tomato yellow leaf curl Sardinia virus is involved in transcriptional repression and severely impacts resistance in Rep-expressing plants. *J Gen Virol* 92:204–209.
128. Mileyko Y, Joh RI, Weitz JS. 2008. Small-scale copy number variation and large-scale changes in gene expression. *Proc Natl Acad Sci* 105:16659–16664.
129. Paprotka T, Deuschle K, Metzler V, Jeske H. 2011. Conformation-Selective Methylation of Geminivirus DNA. *J Virol* 85:12001–12012.
130. Piedra-Aguilera Á, Jiao C, Luna AP, Villanueva F, Dabad M, Esteve-Codina A, Díaz-Pendón JA, Fei Z, Bejarano ER, Castillo AG. 2019. Integrated single-base resolution maps of transcriptome, sRNAome and methylome of Tomato yellow leaf curl virus (TYLCV) in tomato. *Sci Rep* 9:2863.
131. Choi I-R, Stenger DC. 1996. The Strain-Specific cis-Acting Element of Beet Curly Top Geminivirus DNA Replication Maps to the Directly Repeated Motif of the Ori. *Virology* 226:122–126.
132. Behjatnia S, Dry, Rezaian. 1998. Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. *Nucleic Acids Res* 26:925–931.

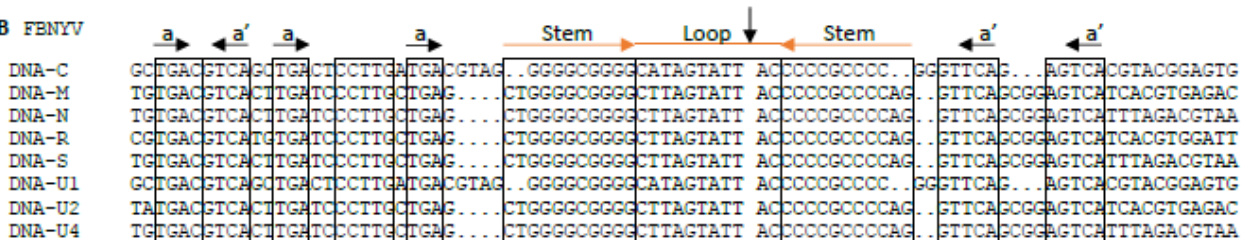
133. Jupin I, Hericourt F, Benz B, Gronenborn B. 1995. DNA replication specificity of TYLCV geminivirus is mediated by the amino-terminal 116 amino acids of the Rep protein. *FEBS Lett* 362:116–120.
134. Ramos PL, Guevara-Gonzalez RG, Peral R, Ascencio-Ibanez JT, Polston JE, Arguello-Astorga GR, Vega-Arreguin JC, Rivera-Bustamante RF. 2003. Tomato mottle Taino virus pseudorecombines with PYMV but not with ToMoV: Implications for the delimitation of cis - and trans -acting replication specificity determinants. *Arch Virol* 148:1697–1712.
135. Gladfelter HJ, Eagle PA, Fontes EPB, Batts L, Hanley-Bowdoin L. 1997. Two Domains of the AL1 Protein Mediate Geminivirus Origin Recognition. *Virology* 239:186–197.
136. Sanz-Burgos AP, Gutiérrez C. 1998. Organization of the cis-Acting Element Required for Wheat Dwarf Geminivirus DNA Replication and Visualization of a Rep Protein–DNA Complex. *Virology* 243:119–129.
137. Chatterji A, Chatterji U, Beachy RN, Fauquet CM. 2000. Sequence Parameters That Determine Specificity of Binding of the Replication-Associated Protein to Its Cognate Site in Two Strains of Tomato Leaf Curl Virus–New Delhi. *Virology* 273:341–350.
138. Revington GN, Sunter G, Bisaro DM. 1989. DNA sequences essential for replication of the B genome component of tomato golden mosaic virus. *Plant Cell* 1:985–992.
139. Schneider M, Jarchow E, Hohn B. 1992. Mutational analysis of the “conserved region” of maize streak virus suggests its involvement in replication. *Plant Mol Biol* 19:601–610.
140. Heyraud F, Matzeit V, Kammann M, Schaefer S, Schell J, Gronenborn B. 1993. Identification of the initiation sequence for viral-strand DNA synthesis of wheat dwarf virus. *EMBO J* 12:4445–4452.
141. 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 Res* 23:910–916.
142. Heyraud F, Matzeit V, Schaefer S, Schell J, Gronenborn B. 1993. The conserved nonanucleotide motif of the geminivirus stem-loop sequence promotes replicational release of virus molecules from redundant copies. *Biochimie* 75:605–615.
143. Stanley J. 1995. Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs within the conserved nonanucleotide motif during the initiation of rolling circle DNA replication. *Virology* 206:707–712.

Supplemental Figure

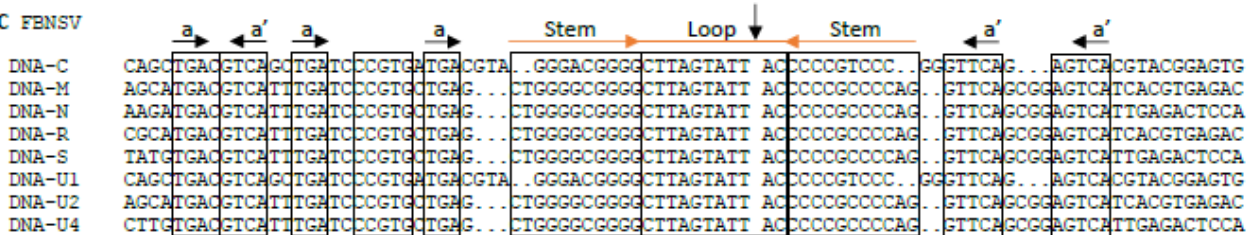
A



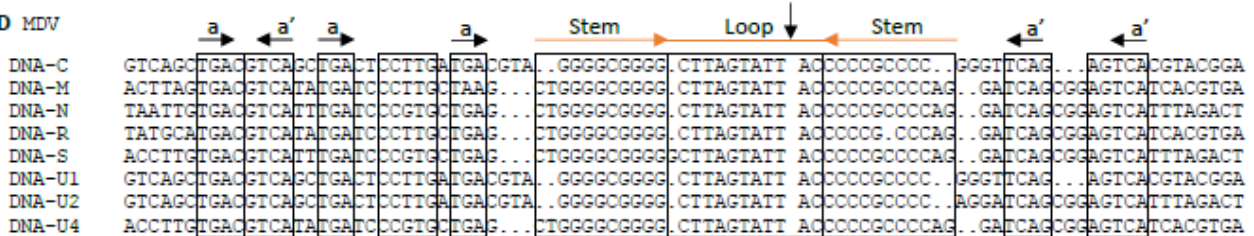
B FBNYV



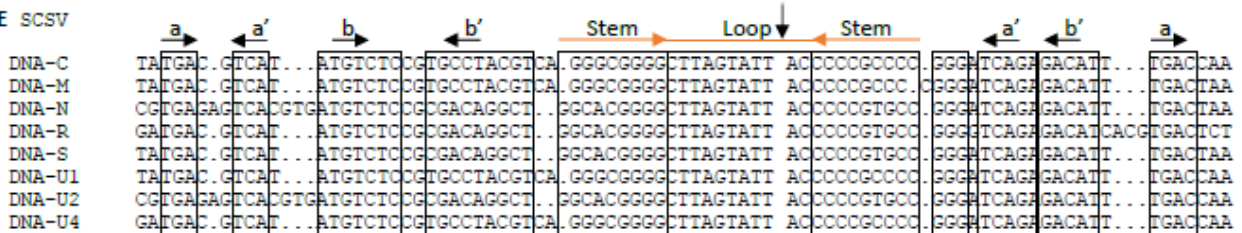
C FBNSV



D MDV



E SCSV



F BBTV

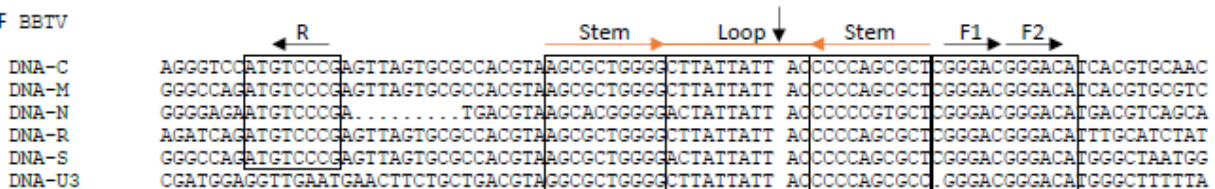


Fig. S1: Alignment of iterative sequences of geminiviruses (A) and nanovirids (B-F). For geminiviruses, TYLCV-S (genus *Begomovirus*, monopartite) BGMV (genus *Begomovirus*, bipartite) and MSV (genus *Mastrevirus*) sequence alignments are adapted from (67). For the nanovirid species FBNYV (B), FBNSV (C), MDV (D), SCSV (E) and BBTV (F), genome component designations are shown on the left. Sequence alignments are adapted from (96–98, 121). Inverted repeat sequences (orange horizontal arrows) forming a stem-loop structure are boxed. Vertical arrows indicate the position of cleavage by the Rep protein within the loop. Conserved sequences shared by genome components are also boxed, revealing several putative iterons (named a, a', b, b', R, F1 and F2 for nanovirids). For SCSV, there are two different groups of sequences for iteron b' on the left side of the stem-loop. Black horizontal arrows indicate iteron sequences.

Chapitre 2 : Etude des mécanismes régulant la formule génomique

I. Contexte

Deux hypothèses sur les mécanismes de régulation de la formule génomique sont proposées dans la littérature : une régulation au niveau de chaque segment ou une régulation au niveau du groupe de segment (Gutiérrez et Zwart, 2018 ; Sicard et al., 2016, 2013 ; Zwart et Elena, 2015). Ainsi, nous avons testé ces hypothèses par différentes approches, parfois redondantes et dont certaines présentaient des incertitudes sur plan conceptuel. Les approches qui ont abouti font l'objet d'un article scientifique qui sera très prochainement soumis et qui est présenté dans la section II. Dans cet article, nous testons d'abord l'éventualité que la formule génomique résulte de capacités d'accumulation (réplication/dégradation) différente selon les segments. Puis, nous nous intéressons à un autre niveau de régulation en considérant le groupe de segments lors de l'infection systémique de la plante.

Dans les sections III et IV de ce chapitre, d'autres approches qui ont été initiées mais n'ont pas abouties sont présentées. Pour chaque approche, le concept, la méthode employée et les perspectives sont discutés.

II. Article : The genome formula of a multipartite virus is regulated both at the individual segment and the segment group levels

Mélia Bonnamy^{1,2}, Andy Brousse^{1,2}, Yannis Michalakis^{2*}, Stéphane Blanc^{1*}

¹ PHIM, Univ Montpellier, IRD, CIRAD, INRAE, Institut Agro, Montpellier, France

² MIVEGEC, CNRS, IRD, Univ Montpellier, Montpellier, France

* These authors contributed equally to this work

Corresponding authors:

Stéphane Blanc (stephane.blanc@inrae.fr), Yannis Michalakis (yannis.michalakis@ird.fr)

Abstract

Differential accumulation of the distinct genome segments is a common feature of viruses with segmented genomes. The reproducible and specific pattern of genome segment accumulation within the host is referred to as the “genome formula”. While there is speculation and some experimental support of a functional role of the genome formula, the mechanisms of its regulation have not yet been identified. In this study, we used a leaf infiltration system to show that the two most accumulated genome segments of the faba bean necrotic stunt virus (FBNSV) possess a greater intrinsic accumulation capacity in *Vicia faba* tissues than the other segments. Nevertheless, processes acting at the individual segment level are insufficient to generate the genome formula, suggesting the involvement of additional mechanisms acting at the supra-segment level. Consistently, the segments accumulation when one segment is absent during a systemic infection indicates that the genome formula is a property of the segment group. Together, these results demonstrate that the FBNSV genome formula is shaped by a complex process acting at both the individual segment and the segment group levels.

Introduction

Viruses can be monopartite, if their entire genetic information is carried by a single nucleic acid molecule; segmented, if their genomic information is carried by several molecules all encapsidated together, or multipartite, if their genomic information is carried by several genome segments which are packaged separately. Hence, in the first two types, the complete viral genome is packaged in a single virion while for multipartite viruses it is distributed over several virus particles.

A remarkable feature of segmented and multipartite viruses is that the fragmented nature of their genome enables the unequal accumulation of the different genome segments in infected tissues. This was first formalized for the multipartite single-stranded DNA (ssDNA) faba bean necrotic stunt virus (FBNSV), where some segments accumulate a lot while others are relatively rare [1]. Subsequently, the differential accumulation of genome segments in the host has been characterized for other multipartite DNA (Bashir et al., 2022; Guyot et al., 2022; Hu et al., 2016; Mansourpour et al., 2022; Yu et al., 2019) and RNA [7] viruses and for segmented RNA viruses [8,9]. Because it was found in viruses with different types of genomes and infecting plants or animals, the differential accumulation of segments seems to be a common feature of viruses with segmented genomes. The pattern of the frequency distribution of the distinct genome segments accumulation within the virus population was referred to as the genome formula [1]. Whenever investigated, the genome formula was shown to be host-specific [1,7,9,10] and the reproducible relative frequency distribution specific to a given host species was defined as the “set-point genome formula” [1].

Several authors speculated that the genome formula has a functional role in the regulation of viral gene expression by differentially modulating their respective copy numbers [1,11–13]. Similar processes, termed amplification-mediated gene expression tuning (AMGET) [14] or genomic accordion [15], were described for procaryotic and eukaryotic microorganisms, and large double stranded monopartite DNA viruses, where rapid adaptation of gene expression in new environments is primarily driven by the selection of copy-number variants [14–16]. Consistently, our group recently showed that the accumulation of FBNSV messenger RNAs is positively correlated to the accumulation of the corresponding genome segments, confirming that genome formula variations modulate gene expression [10].

Upon host-switching, the immediate and reversible host related changes of FBNSV genome formula are not induced by sequence modification [10]. At present, the mechanisms leading to the

reproducible differential accumulation of the genome segments in a given host remain completely unknown. The genome formula could result from selection pressures acting on each segment independently, such as a difference in replication rates across segments [1]. Alternatively, or additionally, the establishment of the genome formula could be driven by selection acting at the level of the group of segments. It has been proposed that groups of segments with different frequency patterns could initiate distinct infection sites at early stages of host colonization, and those with the most optimal frequency pattern for the viral system would propagate faster and thereby be positively selected [1,11,12,17]. This is supported by studies on RNA viruses showing that the very small number of genomes infecting a new cell may generate infection foci with different genetic compositions, increasing the potential for within-host between-foci selection. Adaptive genomes would then be selected, expand the viral infection and constitute the whole virus population [18,19]. In the case of viruses with fragmented genomes, each infection focus with a distinct genome formula could represent a copy number variant and the “set point genome formula” would result from selection of the better fit variant(s).

In this study, we used the FBNSV (family *Nanoviridae*, genus *Nanovirus*) to address the specific question of whether the genome formula is regulated by mechanisms acting on each individual segment or on a higher level involving more than one segment. FBNSV is a phloem-restricted phytovirus belonging to the genus of multipartite viruses with the highest number of genome segments known to date. Its genome is divided into eight circular ssDNA molecules (segments C, M, N, R, S, U1, U2, U4), each encapsidated separately [20]. All segments are assumed to replicate via rolling-circle replication (RCR) controlled by a viral protein of the Rep family, and they all possess a highly conserved replication origin [21,22] (Bonnamy et al., 2023). Each segment is about 1 kb and encodes a single gene. C encodes the protein Clink which interacts with cell cycle regulators to enhance replication [23]. M encodes the movement protein (MP) [24]. N encodes a protein named “nuclear shuttle protein” (NSP) which was proved to be mandatory for aphid transmission [25,26]. R encodes the M-Rep protein which controls the RCR of all genome segments [27,28]. S encodes the capsid protein (CP) which encapsidates each segment individually [29,30]. U1, U2 and U4 encode proteins with unknown functions. The eight FBNSV segments are always found in field collected samples, but it is possible to create incomplete infections in experimental conditions by omitting one or two of the “dispensable” segments [26]. The impact of their absence on the infection phenotype depends on which segment is omitted. In *V. faba*, the absence of either C, N, or U4 does not impact the disease phenotype, although the absence of C reduces the infection

rates by *Agrobacterium*-mediated inoculation. The absence of either U1 or U2 alters the timing and severity of symptoms development. M, R and S are essential.

Using a system of leaf infiltration with pairs of segments, each including R plus one of the seven others, we could estimate the local accumulation of individual segments out of the full genome context and in the absence of systemic movement. This showed that the intrinsic capacity to accumulate locally differs across segments, although segments do not accumulate at ratios that reproduce the systemic genome formula. Thus, processes at the segment level are likely acting, but are not sufficient to explain the genome formula found in systemically infected plants. Infiltrating leaves with all eight segments together did not locally reproduce the genome formula either, further suggesting that other processes occurring during systemic invasion are required. Finally, by deliberately omitting one or two dispensable segments in incomplete infections and analyzing the effect of their absence on the systemic accumulation of the others, we revealed that a group dynamic is also necessary to shape the set-point genome formula.

Results

The intrinsic capacity of the segments to accumulate in plant cells does not reproduce the genome formula

We first investigated whether the genome formula could result from differences in the intrinsic ability of each segment to accumulate in infected plant cells. To test this hypothesis, we quantified the accumulation of each FBNSV genome segment locally, *i.e.* in the absence of systemic movement, and independently of the other segments except for segment R which encodes the protein M-Rep that is mandatory for replication [28]. Fully developed *V. faba* leaves were agro-infiltrated with copies of one of the seven segments C, M, N, S, U1, U2 and U4 along with segment R. The accumulation of each segment in infiltrated tissues was quantified by qPCR six days later. All segments were replicated when co-infiltrated with R, as verified with a replication-threshold estimate described in the Methods section, indicating that R is necessary and sufficient for the replication of all FBNSV segments.

We then compared the relative accumulation of each segment in infiltrated leaves to their respective frequency in systemic infections with the complete FBNSV genome. For this purpose, we calculated the ratio of the accumulation of each segment relative to the accumulation of R in both

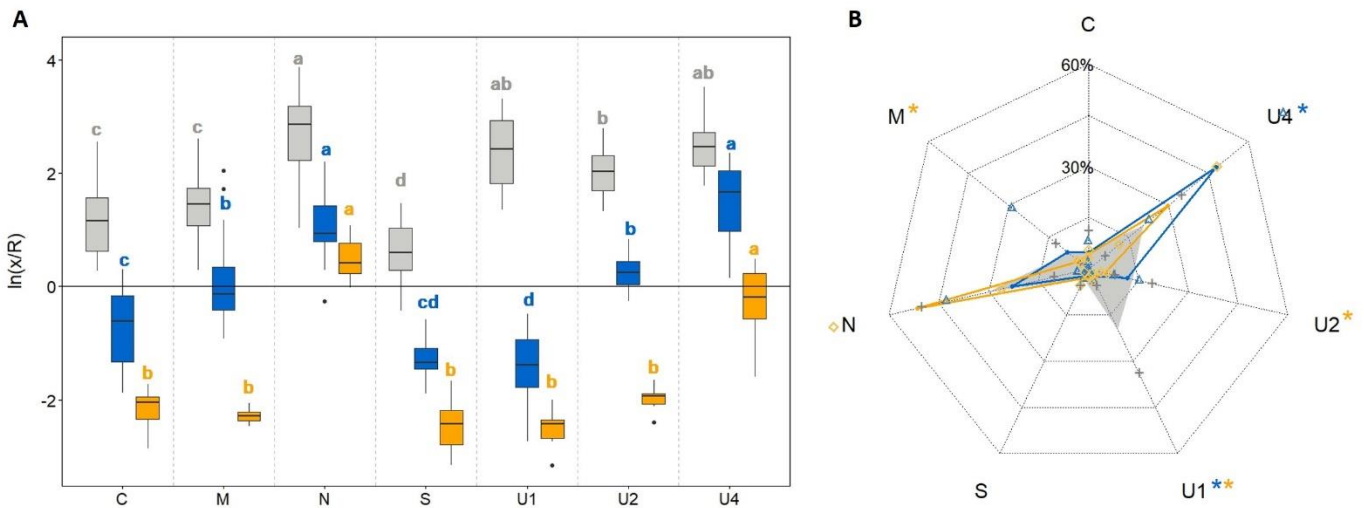


Figure 1: FBNSV segment accumulation in systemically infected or infiltrated leaves of *V. faba*. *V. faba* young plants were stem-agro-inoculated with the eight segments of the FBNSV. DNA was extracted from systemically infected newly formed leaves 4 weeks later (grey). Alternatively, *V. faba* leaves were agro-infiltrated with DNA-R and one of the other seven segments (blue) or with the eight segments (orange). DNA was extracted from infiltrated areas six days later. Virus DNA accumulation in infiltrated and systemically infected leaves was estimated by qPCR. Leaf infiltration data were obtained from two different experiments.

(A) The relative accumulation ratios of each segment with respect to DNA-R was calculated. For each box, the central bar represents the median and the edges of the rectangle the first and third quartiles. The outer bars delineate 1.5 times the distance between the 25th and 75th percentile of the distribution. The dots represent outliers. For a given condition (infiltration in pairs, infiltration of all segments or systemic infection) the differences between segments have been assessed with ANOVA ($p \leq 0.05$) and Tukey HSD tests and are indicated by different letters in grey (systemic infections), blue or yellow (infiltrations).

(B) Relative frequencies of the segments in samples shown in (A): systemic infection (grey), infiltrated with R and another segment (blue) and infiltrated with the eight segments (orange). Standard deviations are represented by grey crosses (systemic infections), blue triangles (infiltration with R) or orange squares (infiltrations with the eight segments). Asterisks associated to segments' name indicate when the differences in frequencies between systemic infections and infiltrations in pairs (blue) or infiltration of all segments (orange) are statistically significant. Differences have been assessed with ANOVA ($p \leq 0.05$) and Tukey HSD tests.

infiltrated and systemically infected leaves (Figure 1A). The normalization of the estimated concentration of each segment by that of R also permitted to account for possible variations of the infiltration efficiency, *i.e.* of the number of transfected cells and thus the total amount of replicated segments in distinct biological replicates. As expected from the earlier characterized genome formula [1], in complete systemic infections, the accumulation ratios of the different segments differ greatly (Figure 1, grey boxes; Table Sup. 1). In infiltrated leaves, the ratios also differ across segments (Figure 1, blue boxes; Table Sup. 2), showing that the genome segments accumulate unequally relatively to R in infiltrated tissues. For all segments, the ratios in infiltrated leaves are lower than those in systemic infections, indicating that R accumulates relatively more in the former condition. We believe this can be explained by a trivial bias of the infiltration method, as further commented in the Discussion section. More importantly, the patterns of the segment ratios are different in infiltrated and in systemically infected leaves. In both infiltrated and systemically infected tissues, N and U4 are the most accumulated segments while S is the least. However, patterns differ for the other segments, the most striking being U1 whose relative accumulation switches from very low in infiltrated leaves to very high in systemic infection.

To better visualize and compare the segment accumulation patterns between infiltrated and systemically infected leaves, we reconstructed a genome formula from the leaf infiltration results: for each segment we calculated the median of the ratios obtained for each sample divided by the sum of the medians of the ratios obtained for each segment (Figure 1B; Table Sup. 3). The accumulation pattern in agro-infiltrated leaves (blue) is different from that of systemically infected leaves (grey), confirming that individual segment accumulation in infiltrated cells does not reproduce the genome formula that develops upon systemic infection of the host plant by the complete FBNSV genome. To determine whether this difference is due to our experimental design where the local accumulation of each segment in infiltrated tissues is considered independently of the other segments, we co-infiltrated the eight FBNSV genome segments together in the same leaves. Unfortunately, for technical reasons related to the agroinfiltration of a mixture of eight distinct bacterial cultures (see Materials and Methods), we obtained only a small number of leaves (five out of twenty infiltrated leaves) with all eight segments amplified by qPCR. In these few leaves, the accumulation pattern proved different both from those infiltrated by single segments and R and from the systemically infected plants (Figure 1; orange; Table Sup. 3; Table Sup. 4). Again, N and U4 are the most accumulated segments, but the six other segments show very low accumulation relative to R.

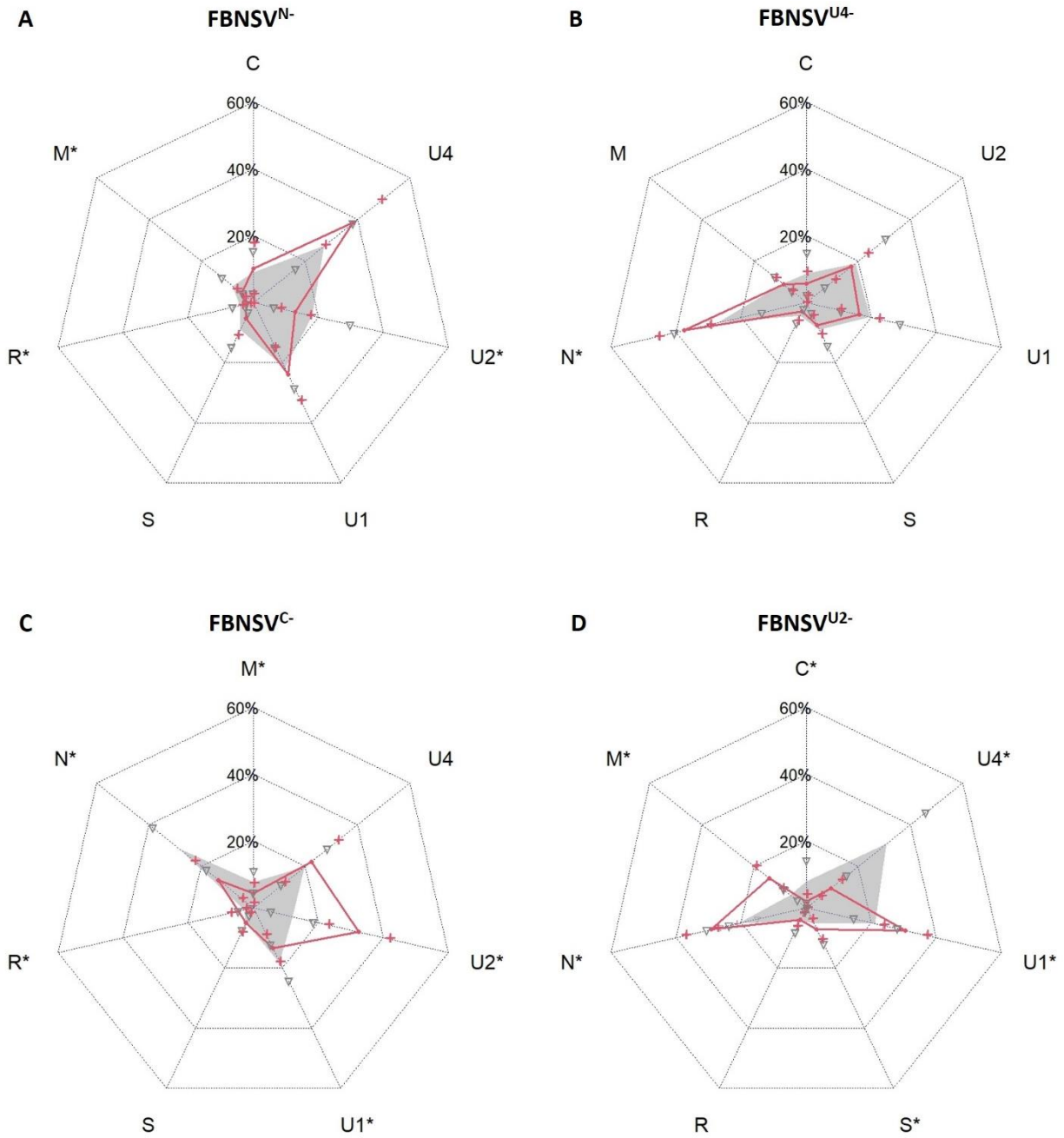


Figure 2: Comparison of the FBNSV genome formula in complete and incomplete infections in *V. faba*. FBNSV genome formula in complete infections (FBNSV^{complete}) is shown in grey whereas that in incomplete infections is shown in red for: (A) FBNSV^{N-}, (B) FBNSV^{U4-}, (C) FBNSV^{C-} or (D) FBNSV^{U2-}. Genome segments accumulation in symptomatic plants has been estimated by qPCR and the relative frequency of the segments was determined. To allow meaningful comparisons, for each radar plot the relative frequency of each segment in complete infections has been calculated without considering the accumulation of the segment which is omitted in incomplete infections. Standard deviations are represented by grey triangles (complete infections) or red crosses (incomplete infections). Asterisks associated to segments' name indicate when the differences in frequencies between complete and incomplete infections are statistically significant (ANOVA ($p \leq 0.05$) and Tukey HSD tests).

In summary, the leaf infiltration experiments show that N and U4, which are the most frequent segments in systemic infections, also accumulate at higher levels locally and individually. Thus, the intrinsic capacity of these segments to accumulate appears to contribute to the genome formula. However, segment accumulation in infiltrated leaves does not entirely reproduce the genome formula, indicating the involvement of a regulatory mechanism occurring at another scale. We thus investigated the segment accumulation pattern at the whole plant scale in the next section.

The genome formula is shaped by mechanisms acting on the segment group during systemic infections

Since the basic capacity of the segments to accumulate locally (individually or together) is not sufficient to explain the genome formula, we investigated their respective accumulation at higher scales, by considering the group of segments and systemic infections. We exploited the possibility to produce incomplete infections [26] by inoculating *V. faba* plantlets with only seven, and in one case six segments. Total DNA was then extracted from systemically infected leaves and the genome formula was determined by calculating the frequency of each segment within the total virus population [1]. We chose to test this in two *a priori* different modalities: i) absence of a segment that does not affect infectivity by agro-inoculation and/or the timing and intensity of symptoms, and ii) absence of a segment that affects at least one of the previously cited traits [26].

Consistent with an earlier report [26], plants infected with FBNSV by all segments except segment N (FBNSV^{N-}) or U4 (FBNSV^{U4-}) exhibited similar time of symptom appearance and similar symptom severity as those infected with all segments (FBNSV^{complete}) (Table Sup. 5). The genome formula of FBNSV^{N-} differed statistically significantly from that of FBNSV^{complete} for segments M, S and U2 even though the mean differences were relatively small (Figure 2A; Table Sup. 6). In plants infected with FBNSV^{U4-}, the accumulation of N was significantly increased. Nevertheless, for both conditions, the segment accumulation pattern is qualitatively similar to that of FBNSV^{complete} (Figure 2B; Table Sup. 6). Hence, the absence of segments N or U4 does not have a qualitative impact on the accumulation of the other segments in our experimental conditions.

Again, as earlier reported [26], the omission of segment C (FBNSV^{C-}) and even more of U2 (FBNSV^{U2-}) led to a decrease of the infection rate in agro-inoculated *V. faba* plants (Table Sup. 5). FBNSV^{C-} showed no differences in the timing of appearance and in the severity of symptoms compared to FBNSV^{complete}, while symptoms were delayed and attenuated for plants infected with FBNSV^{U2-} (Figure Sup. 1; Table Sup. 5). The genome formulas of both FBNSV^{C-} and FBNSV^{U2-} were modified compared to

FBNSV^{complete}. In comparison to plants infected with FBNSV^{complete} all segments but S and U4 were statistically significantly impacted in plants infected with FBNSV^{C-} (Figure 2C; Table Sup. 6), and all segments but R were statistically significantly affected by the absence of U2 (Figure 2D; Table Sup. 6). Hence, the absence of C or U2 differentially modifies the accumulation of the other segments quantitatively and qualitatively, and their relative accumulation pattern depends on the identity of the missing segment.

We also infected plants omitting both C and U4 segments to see the effect on genome formula of the concomitant absence of a segment which does not modify the genome formula (U4) and a segment which does (C). Plants infected with FBNSV^{C-, U4-} showed similar time of symptom appearance and symptom severity compared to FBNSV^{complete}. Interestingly, the genome formula in plants infected with FBNSV^{C-, U4-} were statistically significantly different from that in plants infected with FBNSV^{C-} for S, U1 and U4, but the pattern of accumulation is qualitatively conserved (Figure Sup. 2B; Table Sup. 6), confirming a low impact of U4 on the relative accumulation of other segments.

Overall, our results demonstrate that the absence of a segment differentially affects the frequency of the others. The remaining group of segments apparently adopts a novel set-point genome formula upon systemic infection. Interestingly, this phenomenon does not occur when the missing segment is N or U4, which have no impact on the disease development, but are highly accumulated during *V. faba* infection.

Discussion

We first determined that segment R is sufficient for the replication of any FBNSV genome segment in agro-infiltrated *V. faba* leaves. This was expected since Timchenko and collaborators demonstrated that pairwise infiltrations of each segment of the nanovirus faba bean necrotic yellows virus (FBNYV) with R is sufficient to induce their replication in *Nicotiana benthamiana* leaf discs [27,28]. We thus verified here that, also for FBNSV, R is the only segment mandatory for the replication of all genome segments *in planta*.

More surprising was the observation that, relative to the co-infiltrated segment R, FBNSV segments do not accumulate equally. These differences between individual segments could result from segment-specific replication rates. Considering that all segments have approximately the same size, different replication rates should be due to distinct regulatory processes. Two elements are thought to be

of primary importance for nanovirus replication: the iterons which determine the recognition of viral DNA by the Rep protein, and a sequence that forms a stem-loop structure and constitutes the replication origin. The FBNSV iteron sequences are identical for the eight segments. The stem sequence is identical for segments C/U1 and for segments M/N/R/S/U2/U4, but slightly differs between these two groups to which it may confer a different stem-loop stability. However, these sequence variations are not consistent with the results obtained from the leaf infiltration experiments because segments of the same stem sequence group (for example S and U4) differ greatly in their accumulation levels. Differential replication rates could be caused by other yet unidentified regulatory sequences or by differences in the synthesis of the viral complementary strand which serves as the template for replication (Bonnamy et al., 2023). Alternatively, accumulation differences between segments could result from differential degradation of the DNA segments due to the nature of their sequence and secondary structures.

When taken individually, though the distinct segments accumulate at different concentrations, their normalized relative accumulation pattern is not consistent with that observed when they systemically infect hosts together. First, R is much more accumulated than other segments in infiltrated compared to systemically infected leaves (Figure 1A). Two characteristics of plant systemic infections by the FBNSV must be considered to tentatively interpret this observation: FBNSV is naturally restricted to the companion cells [31], and its genome segments rarely co-localize in individual cells which implies a cross complementation by inter-cellular movement of gene expression products [32]. In leaf infiltration experiments, the infiltrated cell types are not identified but are most likely mesophyll cells. Obviously, FBNSV can replicate in this type of cells, but it is likely that there are no intercellular movements of viral products (DNA, RNA or proteins) as these are naturally moving in and out solely of sieve elements and companion cells. A series of three studies in the genera *Babuvirus* and *Nanovirus* on the tissue specificity of the promotor of the segment R [33–35] indicated some degree of specificity for the vascular cells. In our case, however, because the simultaneous co-infiltration of the eight FBNSV segments in fully developed leaves never led to systemic infections (none out of 50 infiltrated plants; data not shown), it seems unlikely that the infiltrated segments often reached the vascular bundles. The consequence of this artificial leaf-infiltration system is thus that complementation between segments may be possible only if they penetrate the same cell. Any segment penetrating a cell in the absence of R would not be able to replicate, while R alone can. Thus, the relative accumulation of R measured in infiltrated leaves may well be overestimated compared to that of the other segments and most likely does not reflect an intrinsically higher accumulation. Despite this obvious possibility, the other segments inoculated in pairs with R should

a priori all have the same chance to penetrate a cell in which R is present and so their accumulation should be comparable. When comparing them, their accumulation pattern does not match that observed upon systemic infection (Figure 1). This could have several explanations: (i) the infected cell type, (ii) the absence of interaction between the different segments beyond that with R, or (iii) the local scale precluding successive replication cycles when colonizing newly formed tissues.

Relative to the first potential explanation, in infiltrated leaves the FBNSV segments are mainly replicated in mesophyll cells, and perhaps in a few epidermic cells, while during systemic infections the virus is phloem-limited. It is possible that the cell type impacts the relative accumulation of segments. However, as stated above and due to near identical regulatory sequences, it seems unlikely that the cell type can impact differentially the accumulation of each segment. To our knowledge, this has not been investigated and we do not see an immediate feasible approach to test for it.

With respect to the second potential explanation, by co-infiltrating each segment solely with R we were interested in the ability of each segment to accumulate independently of the others. Nevertheless, when together, interactions between segments could impact their accumulation. For example, there could be competition between segments for replication, or the expression product of a given segment could affect the accumulation of another (including R). Viral DNA-protein or protein-protein interactions have been identified in some members of the family *Nanoviridae* [24,36–38], but their relevance for the replication of different genome segments has not been investigated. To study whether interactions between segments (or their expression products) could affect their accumulation locally, we co-infiltrated the eight FBNSV genome segments together in the same leaf. The pattern of segment accumulation when the eight segments are infiltrated together is different from that when infiltrated solely with R, as well as in systemically infected leaves (Figure 1). N and U4 appear highly accumulated in all three experimental conditions, which may suggest that they are more competent for replication. However, the interpretation of the results obtained with co-infiltration of the eight segments should be very cautious. Indeed, it is likely that each cell does not necessarily internalize all segments together, and the probable lack of movement of viral material across distinct cells makes the interactions between the segments (or their expression products) uncontrollable. Therefore, apart from the high accumulation of U4 and N, we cannot draw further sound conclusions on the impact of inter-segment interactions on their local accumulation.

As for the third potential explanation, the leaf infiltration system allows to study segment accumulation individually, and probably at a cellular level as discussed above. However, the genome

formula could result from mechanisms operating at a supra-cellular level. Our group [39] earlier estimated that the effective population sizes of two FBNSV segments (S and N) during transmission by the vector and plant colonization is relatively low. The low number of genomic segments propagating within hosts and initiating new infection foci could generate random variation of the frequency of genome segments across foci. This phenomenon designated genome formula drift [11,13] could enhance selection at the between foci level [18,19] on the genome formula. During successive cycles of infection of new cells or tissues, the infection foci that would multiply the most would be those for which the genome formula is closer to the optimal, rapidly driving the virus population towards the set-point genome formula [1,11,13]. Would this hypothesis be valid, then the set-point genome formula could be different from the “local genome formula”, where no successive selection cycles occur, and where the formula is determined solely by the capacity of each segment to accumulate during one replication round. In addition, the viral system as a whole would adapt to environmental changes by adopting a new optimal formula.

To further investigate segment accumulation at the systemic level, we determined the set-point genome formulas in incomplete infections. The absence of a segment that has no impact on the infection (N or U4) does not qualitatively affect the relative frequencies of the remaining segments. Conversely, when the absence of a segment negatively affects FBNSV infection of *V. faba* (C or U2), this absence qualitatively impacts the accumulation of other segments that must collectively adjust to a novel set-point formula (Figure 2). These conclusions support the hypothesis of a selection at the level of the segment group [1,11,17].

The absence of N or U4 does not impact the accumulation of the other segments, and interestingly, they are also distinguished from other segments by the fact that they have a greater intrinsic capacity to accumulate in *V. faba* infiltrated cells. N is required for aphid vector transmission [25,26], but its potential involvement in other steps of the viral cycle is unclear. It has been proposed that the NSP of the babuvirus BBTV may assist in the transport of viral proteins or DNA out of the nucleus [36], but its absence does not appear to impact FBNSV infection *in planta*, and its function during transmission may be effective solely during the crossing of aphid vector gut or salivary gland cells [25,40]. The function of the U4 protein is entirely unknown, and it has been reported repeatedly to have no within-host function, at least in experimental conditions. If these two segments are totally dispensable *in planta* and have even no impact on the genome formula, and therefore on the other segments, they could potentially get lost upon successive bottlenecks during progression into uninfected and newly formed tissues. Because in all the experiments reported in this study they both consistently appear to be highly accumulated, we could

speculate that they have unknown regulatory sequences which make them more replication/accumulation competent than the others. This could have been selected to avoid their loss since at least N is mandatory to complete the natural life cycle.

In conclusion, the genome formula of FBNSV seems to be determined for one part by regulation at the segment level, especially for the very frequent N and U4, and for another part by processes acting at the supra-segment level. Our results thus support the hypothesis that the unit of selection for the genome formula is at a supra-segment level.

Materials and methods

Plants, growth conditions, and FBNSV infectious clone

Faba beans (*Vicia faba*) var. Seville (Vilmorin) were grown in a growth chamber with 13h/11h day/night photoperiod, 25/18°C day/night temperature.

The FBNSV infectious isolate JKI-2000 used in this study was cloned by [20]. The infectious clone is constructed as eight binary plasmids (derived from pBIN-19), each containing a tandem repeat of one of the segments. Each of these plasmids was used to transform one colony of *Agrobacterium tumefaciens* strain COR308, yielding eight bacterial clones each capable of transferring one of the eight segments (C, M, N, R, S, U1, U2, U4) to a host plant upon agro-infiltration.

Leaf agro-infiltrations

The eight *A. tumefaciens* clones were separately grown in NZY+ medium (0.1% NZ amine, 0.5% yeast extract, 0.5% NaCl at pH7.5) completed with 12.5 mM MgCl₂, 12.5 mM MgSO₄ and 0.4% glucose, 10 mM MES pH 5.5, 50 μM acetosyringone, 50 mg.ml⁻¹ kanamycin, 25 mg.ml⁻¹ gentamycin and 5 mg.ml⁻¹ tetracycline, at 28°C with stirring at 150 rpm for 16 h. Bacteria were then pelleted by centrifugation and resuspended in 10 mM MgCl₂ solution. Different mixes of agrobacteria solutions were prepared depending on the experimental design: each segment alone (Optical Density (OD) = 0.4), mixes of pairs of segments containing R (OD = 0.4) and one of segments C, M, N, S, U1, U2, or U4 (OD = 0.4), or mixes containing R (OD = 0.4) and all the seven other segments (OD = 0.06 for each). Each mix was infiltrated on the underside of 14 days-old *V. faba* plantlets fully expanded leaves.

Systemic infections

Systemic infections were obtained by agro-inoculating *V. faba* plants based on the protocol described earlier [1]. *A. tumefaciens* solutions were prepared as described for leaf agro-infiltration experiments and were inoculated by picking and injecting into the stem of nine days old seedlings. Complete infections (FBNSV^{complete}) were made by inoculating a mix of the eight *A. tumefaciens* clones, re-constituting the full FBNSV infectious clone. For incomplete infections, only seven (FBNSV^C, FBNSV^N, FBNSV^{U2}, FBNSV^{U4}) *Agrobacterium* clones were mixed and inoculated.

For plants inoculated by FBNSV^{U2} only one infected plant was obtained by *Agrobacterium*-mediated inoculation. This plant was used as a source to inoculate other plants by aphids (*Acyrtosiphon pisum*, clone 210). After a three days acquisition access period (AAP) on the infected plant (four weeks post agro-inoculation), viruliferous aphids were transferred on eleven days old *V. faba* plants (ten aphids per plant). After an inoculation access period (IAP) of two days, aphids were collected, mixed, and used to inoculate a second batch of thirteen days old plants for a two-day IAP (eight or nine aphids per plant). After this second inoculation, the two batches of plants were treated with Pirimor G (1 g.L⁻¹ in water) to kill aphids.

Plants infected with FBNSV^{C, U4} were obtained by inoculating eight days old *V. faba* plantlets with one aphid (*A. pisum*) that fed first on FBNSV^C for three days, then on FBNSV^{U4} for three more days or vice versa. After three days of inoculation, plants were treated with Pirimor. Plants in which neither C nor U4 were transmitted were used for this study.

DNA extraction and qPCR conditions

For infiltrated leaves, total DNA was extracted from infiltrated tissues six days after infiltrations following the protocol described by [41]. For systemically infected plants, total DNA was extracted from symptomatic apical leaves three (FBNSV^C and FBNSV^{C, U4}) or four (FBNSV^N, FBNSV^{U2} and FBNSV^{U4}) weeks after inoculation following the same protocol. The concentration of total extracted DNA was estimated using a spectrophotometer NanoDrop™ 2000 (Thermo Scientific™, Waltham, MA, USA) and that of each genome segment was then determined by qPCR using the LightCycler FastStart DNA Master Plus SYBR Green I kit (Roche, Indianapolis, Ind, USA). Following the manufacturer's instructions, 5 µL of the 2X qPCR Mastermix were mixed with segment-specific primers (Table Sup. 6) at 0.3 µM (segments C, M and S) or 0.5 µM (segments N, R, U1, U2, U4) final, 2 µL of 10-fold diluted DNA extracts and complemented with

water to obtain a final reaction volume of 10 μ L. qPCR reactions were carried out in a LightCycler 480 thermocycler (Roche) with 40 cycles of 95°C for 10 s, 60°C for 10 s and 72°C for 10 s. Two technical replicates were done for each sample. Post-PCR data analyses were carried out with the LinRegPCR software [42].

Data analysis

(i) Leaf infiltration experiments

For each sample, the DNA concentration measured by qPCR was normalized by the concentration of total DNA extracted from the leaves. Thresholds were then determined to remove samples where FBNSV segments have not been replicated in presence of segment R. Samples with concentrations below the threshold were excluded from the analysis as we considered that the segment was not replicated.

To estimate the thresholds, we quantified virus DNA segments from ten leaves infiltrated with either of the seven segments (C, N, S, U1, U2, U4) alone, without R. We could thus determine the qPCR basal level measured in the absence of replication (no segment R). We then calculated the upper tolerance threshold for each FBNSV segment such that 95% of potential future samples in which the segment is absent would have lower estimated DNA concentration with a 95% probability, based on the method described by [43,44]. The thresholds were 1.990203e-06 for segment C, 5.402976e-06 for M, 3.799221e-06 for N, 3.068326e-06 for S, 1.709007e-06 for U1, 3.235822e-06 for U2 and 3.716083e-06 for U4.

To compare the accumulation of each DNA segment in agro-infiltrated leaves (in pair with R or with all segments) to their respective accumulation in systemic infections, the concentration of each individual segment was divided by that of R in the same sample. We then used the RStudio software to run linear models on the log-transformed data to make their distribution closer to the Normal distribution. Data were obtained from two independent experimental replicates.

Genome formulas were reconstructed from the results of agro-infiltrations and were compared to the genome formula in systemic infections. For each sample, the accumulation of each segment normalized by that of R was divided by the sum of the medians across all samples of the accumulation of each segment normalized by R.

(ii) Incomplete infections experiment

Plants with one or more missing segments other than those voluntarily omitted were excluded from the analysis. The frequency of each segment was expressed relatively to the total virus DNA for incomplete experiments. To compare the frequencies in incomplete and complete infections, the relative frequency of the segments in complete infections was calculated without considering the accumulation of the segment which is removed in incomplete infections. We then used JMP to run linear models on the logit-transformed data to make their distribution closer to the Normal distribution. Results were obtained from one (FBNSV^{U2-}, FBNSV^{C-, U4-}), two (FBNSV^{N-}, FBNSV^{U4-}) or three (FBNSV^{C-}) independent experimental replicates.

Acknowledgments

This work was funded by the French ANR : Nanovirus ANR-18-CE92-0028, as well as by the University of Montpellier MUSE, project MULTIVIR. MB acknowledges support from University of Montpellier, IRD, CNRS and INRAE. SB acknowledges support from INRAE and Department SPE. YM acknowledges support from CNRS and IRD. We thank technical help provided by Ms Sophie Le Blaye.

References

1. Sicard A, Yvon M, Timchenko T, Gronenborn B, Michalakis Y, Gutierrez S, et al. Gene copy number is differentially regulated in a multipartite virus. *Nat Commun.* 2013;4: 2248. doi:10.1038/ncomms3248
2. Bashir S, Farrakh S, Yasmin T, Muhammad A, Bashir T, Manghwar H, et al. Quantitation of Multipartite Banana Bunchy Top Virus Genomic Components and Their Transcripts in Infected Tissues of Banana (*Musa acuminata*). *Agronomy.* 2022;12: 2990. doi:10.3390/agronomy12122990
3. Guyot V, Rajeswaran R, Chu HC, Karthikeyan C, Laboureau N, Galzi S, et al. A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids. Bisaro DM, editor. *PLOS Pathog.* 2022;18: e1010448. doi:10.1371/journal.ppat.1010448
4. Hu Z, Zhang X, Liu W, Zhou Q, Zhang Q, Li G, et al. Genome segments accumulate with different frequencies in *Bombyx mori bidensovirus*: The different frequencies of BmBDV genome segments. *J Basic Microbiol.* 2016;56: 1338–1343. doi:10.1002/jobm.201600120
5. Mansourpour M, Gallet R, Abbasi A, Blanc S, Dizadji A, Zeddami J-L. Effects of an Alphasatellite on the Life Cycle of the Nanovirus *Faba Bean Necrotic Yellows Virus*. Simon AE, editor. *J Virol.* 2022;96: e01388-21. doi:10.1128/JVI.01388-21
6. Yu N-T, Xie H-M, Zhang Y-L, Wang J-H, Xiong Z, Liu Z-X. Independent modulation of individual genomic component transcription and a cis-acting element related to high transcriptional activity in a multipartite DNA virus. *BMC Genomics.* 2019;20: 573. doi:10.1186/s12864-019-5901-0
7. Wu B, Zwart MP, Sánchez-Navarro JA, Elena SF. Within-host Evolution of Segments Ratio for the Tripartite Genome of Alfalfa Mosaic Virus. *Sci Rep.* 2017;7: 5004. doi:10.1038/s41598-017-05335-8
8. Diefenbacher M, Sun J, Brooke CB. The parts are greater than the whole: the role of semi-infectious particles in influenza A virus biology. *Curr Opin Virol.* 2018;33: 42–46. doi:10.1016/j.coviro.2018.07.002
9. Moreau Y, Gil P, Exbrayat A, Rakotoarivony I, Bréard E, Sailleau C, et al. The Genome Segments of Bluetongue Virus Differ in Copy Number in a Host-Specific Manner. Parrish CR, editor. *J Virol.* 2020;95: e01834-20. doi:10.1128/JVI.01834-20
10. Gallet R, Di Mattia J, Ravel S, Zeddami J-L, Vitalis R, Michalakis Y, et al. Gene copy number variations at the within-host population level modulate gene expression in a multipartite virus. *Virus Evol.* 2022;8: veac058. doi:10.1093/ve/veac058
11. Gutiérrez S, Zwart MP. Population bottlenecks in multicomponent viruses: first forays into the uncharted territory of genome-formula drift. *Curr Opin Virol.* 2018;33: 184–190. doi:10.1016/j.coviro.2018.09.001
12. Sicard A, Michalakis Y, Gutiérrez S, Blanc S. The Strange Lifestyle of Multipartite Viruses. Hobman TC, editor. *PLOS Pathog.* 2016;12: e1005819. doi:10.1371/journal.ppat.1005819
13. Zwart MP, Elena SF. Modeling multipartite virus evolution: the genome formula facilitates rapid adaptation to heterogeneous environments†. *Virus Evol.* 2020;6: veaa022. doi:10.1093/ve/veaa022
14. Tomanek I, Grah R, Lagator M, Andersson AMC, Bollback JP, Tkačik G, et al. Gene amplification as a form of population-level gene expression regulation. *Nat Ecol Evol.* 2020;4: 612–625. doi:10.1038/s41559-020-1132-7

15. Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, et al. Poxviruses Deploy Genomic Accordions to Adapt Rapidly against Host Antiviral Defenses. *Cell*. 2012;150: 831–841. doi:10.1016/j.cell.2012.05.049
16. Bayer A, Brennan G, Geballe AP. Adaptation by copy number variation in monopartite viruses. *Curr Opin Virol*. 2018;33: 7–12. doi:10.1016/j.coviro.2018.07.001
17. Zwart MP, Elena SF. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu Rev Virol*. 2015;2: 161–179. doi:10.1146/annurev-virology-100114-055135
18. Miyashita S, Ishibashi K, Kishino H, Ishikawa M. Viruses Roll the Dice: The Stochastic Behavior of Viral Genome Molecules Accelerates Viral Adaptation at the Cell and Tissue Levels. Sugden B, editor. *PLOS Biol*. 2015;13: e1002094. doi:10.1371/journal.pbio.1002094
19. Miyashita S, Kishino H. Estimation of the Size of Genetic Bottlenecks in Cell-to-Cell Movement of *Soil-Borne Wheat Mosaic Virus* and the Possible Role of the Bottlenecks in Speeding Up Selection of Variations in *trans*-Acting Genes or Elements. *J Virol*. 2010;84: 1828–1837. doi:10.1128/JVI.01890-09
20. Grigoras I, Timchenko T, Katul L, Grande-Pérez A, Vetten H-J, Gronenborn B. Reconstitution of Authentic Nanovirus from Multiple Cloned DNAs. *J Virol*. 2009;83: 10778–10787. doi:10.1128/JVI.01212-09
21. Gronenborn B. Nanoviruses: genome organisation and protein function. *Vet Microbiol*. 2004;98: 103–109. doi:10.1016/j.vetmic.2003.10.015
22. Lal A, Vo TTB, Sanjaya IGNPW, Ho PT, Kim J-K, Kil E-J, et al. Nanovirus Disease Complexes: An Emerging Threat in the Modern Era. *Front Plant Sci*. 2020;11: 558403. doi:10.3389/fpls.2020.558403
23. Aronson MN, Meyer AD, Györgyey J, Katul L, Vetten HJ, Gronenborn B, et al. Clink, a Nanovirus-Encoded Protein, Binds both pRB and SKP1. *J Virol*. 2000;74: 2967–2972. doi:10.1128/JVI.74.7.2967-2972.2000
24. Wanitchakorn R, Hafner GJ, Harding RM, Dale JL. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *Microbiology*. 2000;81: 299–306. doi:10.1099/0022-1317-81-1-299
25. Di Mattia J, Vernerey M-S, Yvon M, Piroilles E, Villegas M, Gaafar Y, et al. Route of a Multipartite Nanovirus across the Body of Its Aphid Vector. Simon AE, editor. *J Virol*. 2020;94: e01998-19. doi:10.1128/JVI.01998-19
26. Grigoras I, Vetten H-J, Commandeur U, Ziebell H, Gronenborn B, Timchenko T. Nanovirus DNA-N encodes a protein mandatory for aphid transmission. *Virology*. 2018;522: 281–291. doi:10.1016/j.virol.2018.07.001
27. Timchenko T, Katul L, Sano Y, de Kouchkovsky F, Vetten HJ, Gronenborn B. The master rep concept in nanovirus replication: identification of missing genome components and potential for natural genetic reassortment. *Virology*. 2000;274: 189–195. doi:10.1006/viro.2000.0439
28. Timchenko T, Kouchkovsky F de, Katul L, David C, Vetten HJ, Gronenborn B. 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*. 1999;73: 10173–10182.
29. Chu PWG, Paul K, Bing-sheng Q, Waterhouse PM, Gerlach WL. Putative full-length clones of the genomic DNA segments of subterranean clover stunt virus and identification of the segment coding for the viral coat protein. *Virus Res*. 1993;27: 161–171. doi:10.1016/0168-1702(93)90079-3
30. Katul L, Maiss E, Morozov SY, Vetten HJ. Analysis of six DNA components of the faba bean necrotic yellows virus genome and their structural affinity to related plant virus genomes. *Virology*. 1997;233: 247–259. doi:10.1006/viro.1997.8611

31. Thomas JE, Gronenborn B, Harding RM, Mandal B, Grigoras I, Randles JW, et al. ICTV Virus Taxonomy Profile: Nanoviridae. *J Gen Virol.* 2021;102. doi:10.1099/jgv.0.001544
32. Sicard A, Pirolles E, Gallet R, Vernerey M-S, Yvon M, Urbino C, et al. A multicellular way of life for a multipartite virus. *eLife.* 2019;8: e43599. doi:10.7554/eLife.43599
33. Dugdale B, Becker DK, Beetham PR, Harding RM, Dale JL. Promoters derived from banana bunchy top virus DNA-1 to -5 direct vascular-associated expression in transgenic banana (*Musa* spp.). *Plant Cell Rep.* 2000;19: 810–814. doi:10.1007/s002999900185
34. Dugdale B, Beetham PR, Becker DK, Harding RM, Dale JL. Promoter activity associated with the intergenic regions of banana bunchy top virus DNA-1 to -6 in transgenic tobacco and banana cells. *J Gen Virol.* 1998;79: 2301–2311. doi:10.1099/0022-1317-79-10-2301
35. Shirasawa-Seo N, Sano Y, Nakamura S, Murakami T, Seo S, Ohashi Y, et al. Characteristics of the promoters derived from the single-stranded DNA components of Milk vetch dwarf virus in transgenic tobacco. *J Gen Virol.* 2005;86: 1851–1860. doi:10.1099/vir.0.80790-0
36. Ji X, Yu N, Qu L, Li B, Liu Z. Banana bunchy top virus (BBTV) nuclear shuttle protein interacts and re-distributes BBTV coat protein in *Nicotiana benthamiana*. *3 Biotech.* 2019;9: 121. doi:10.1007/s13205-019-1656-1
37. Krenz B, Schießl I, Greiner E, Krapp S. Analyses of pea necrotic yellow dwarf virus-encoded proteins. *Virus Genes.* 2017;53: 454–463. doi:10.1007/s11262-017-1439-x
38. Yu N, Wang J, Yu N, Zheng X, Zhou Q, Liu Z. Bioinformatics Analysis of the Interaction between Coat Protein and Nuclear Shuttle Protein in Babuvirus. *Am J Plant Sci.* 2019;10: 622–630. doi:10.4236/ajps.2019.104045
39. Gallet R, Fabre F, Thébaud G, Sofonea MT, Sicard A, Blanc S, et al. Small Bottleneck Size in a Highly Multipartite Virus during a Complete Infection Cycle. Simon AE, editor. *J Virol.* 2018;92: e00139-18. doi:10.1128/JVI.00139-18
40. Di Mattia J, Zeddami J-L, Uzest M, Blanc S. The helper strategy in vector-transmission of plant viruses. *Peer Community J.* 2023;3: e32. doi:10.24072/pcjournal.258
41. Di Mattia J, Ryckebusch F, Vernerey M-S, Pirolles E, Sauvion N, Peterschmitt M, et al. Co-Acquired Nanovirus and Geminivirus Exhibit a Contrasted Localization within Their Common Aphid Vector. *Viruses.* 2020;12: 299. doi:10.3390/v12030299
42. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den Hoff MJB, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* 2009;37: e45–e45. doi:10.1093/nar/gkp045
43. Di Mattia J, Torralba B, Yvon M, Zeddami J-L, Blanc S, Michalakakis Y. Nonconcomitant host-to-host transmission of multipartite virus genome segments may lead to complete genome reconstitution. *Proc Natl Acad Sci.* 2022;119: e2201453119. doi:10.1073/pnas.2201453119
44. Francq BG, Lin D, Hoyer W. Confidence, prediction, and tolerance in linear mixed models. *Stat Med.* 2019;38: 5603–5622. doi:10.1002/sim.8386

Supplemental Figures

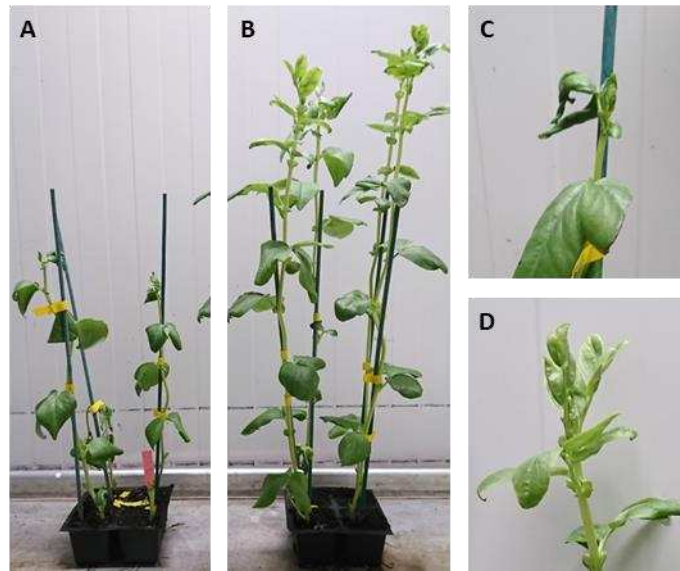


Figure Sup. 1: Symptoms of *V. faba* infected with FBNSV^{complete} (A and C) or FBNSV^{U2} (B and D) four weeks after agro-inoculation

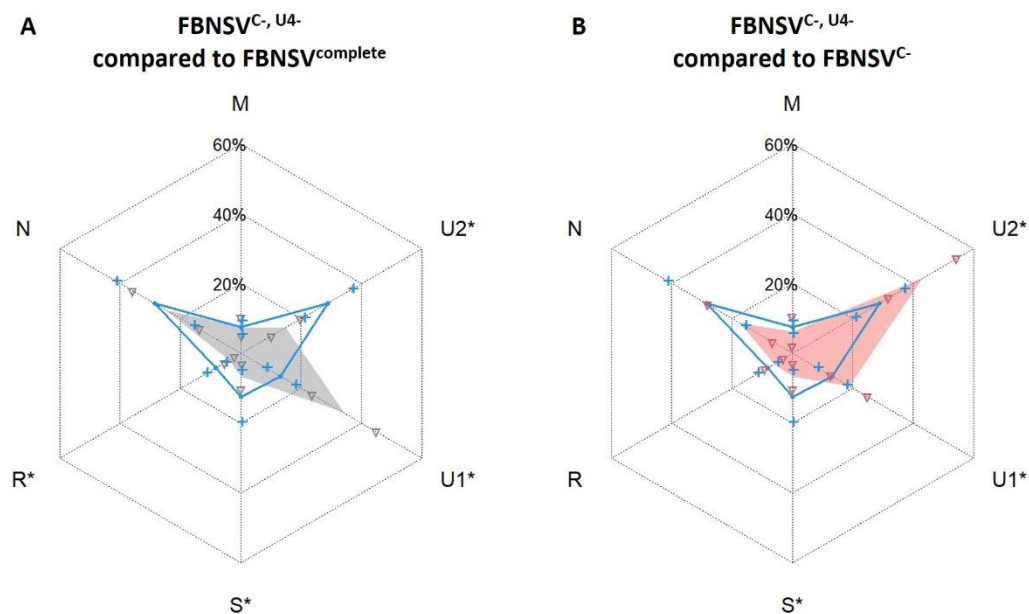


Figure Sup. 2: Comparison of the FBNSV genome formula in complete and incomplete infections without C and/or U4. FBNSV genome formula in incomplete infections FBNSV^{C-, U4-} (blue) compared to FBNSV^{complete} (A; grey) or FBNSV^{C-} (B; red). Genome segments accumulation in symptomatic *V. faba* plants has been estimated by qPCR and the relative frequency of the segments was determined. To allow meaningful comparisons, the relative frequency of each segment has been calculated without considering the accumulation of segments C and U4 in FBNSV^{complete} and without considering U4 in FBNSV^{C-}. Standard deviations are represented by grey triangles (complete infections) or red crosses (incomplete infections). Asterisks associated to segments' name indicate when the differences in frequencies between complete and incomplete infections are statistically significant (ANOVA ($p \leq 0.05$) and Tukey HSD tests).

Supplemental Tables

Table Sup. 1: Statistical analysis of the comparison of segment accumulation relative to R in systemically infected leaves

Comparisons of accumulation relative to R of each segment to that of the others in systemically infected leaves were assessed through ANOVA and Tukey HSD tests on Ln-transformed data and using RStudio. The p-value indicating significant differences (≤ 0.05) are in red.

Segment	C	M	N	S	U1	U2	U4
C							
M	0.9055451						
N	0.0000000	0.0000000					
S	0.0458036	0.0009860	0.0000000				
U1	0.0000002	0.0000398	0.4128860	0.0000000			
U2	0.0006140	0.0322792	0.0034745	0.0000000	0.5257508		
U4	0.0000000	0.0000016	0.8564561	0.0000000	0.9909110	0.1472480	

Table Sup. 2 : Statistical analysis of the comparison of segment accumulation relative to R in leaves infiltrated with each segment in pair with R

Comparisons of accumulation relative to R of each segment to that of the others in leaves infiltrated with pairs of segments were assessed through ANOVA and Tukey HSD tests on Ln-transformed data and using RStudio. The p-value indicating significant differences (≤ 0.05) are in red.

Segment	C	M	N	S	U1	U2	U4
C							
M	0.0006474						
N	0.0000000	0.0021410					
S	0.0937588	0.0000000	0.0000000				
U1	0.0143595	0.0000000	0.0000000	0.9982984			
U2	0.0000200	0.9998172	0.0023321	0.0000000	0.0000000		
U4	0.0000000	0.0000001	0.2003704	0.0000000	0.0000000	0.0000000	

Table Sup. 3: Statistical analysis of the comparison of segment accumulation according to the condition

Comparisons of accumulation relative to R of each segment in each condition (systemic infections, leaf infiltration with pairs of segments or infiltrations with the eight segments) were performed two by two and assessed through ANOVA and Tukey HSD tests on Ln-transformed data and using RStudio. The p-value indicating significant differences after FDR correction (≤ 0.05) are in red.

segment	Systemic infections/infiltrations – pairs		Systemic infections/infiltrations – 8 segments		Infiltrations – pairs/infiltrations – 8 segments	
	p-value	significance	p-value	significance	p-value	significance
C	0.15		0.187		0.745	
M	0.131		0.00231	**	0.00907	**
N	0.861		0.0224	*	0.00999	**
S	0.0572		0.626		0.358	
U1	1.02e-11	***	5.64e-07	***	0.456	
U2	0.343		7.35e-06	***	1.72e-06	***
U4	0.000586	***	0.513		0.117	

Table Sup. 4: Statistical analysis of the comparison of segment accumulation relative to R in leaves infiltrated with the eight FBNSV segments

Comparisons of accumulation relative to R of each segment to that of the others in leaves infiltrated with the eight segments were assessed through ANOVA and Tukey HSD tests on Ln-transformed data and using RStudio. The p-value indicating significant differences (≤ 0.05) are in red.

Segment	C	M	N	S	U1	U2	U4
C							
M	0.9995170						
N	0.0000000	0.0000000					
S	0.9332841	0.9945321	0.0000000				
U1	0.8307446	0.9675890	0.0000000	0.9999653			
U2	0.9915625	0.9167585	0.0000000	0.5755166	0.4102672		
U4	0.0000006	0.0000002	0.0651658	0.0000000	0.0000000	0.0000046	

Table Sup. 5: Infection rates and phenotypes of FBNSV complete and incomplete infections in *V. faba*

Infections rates and symptom severity were determined three (FBNSV^{complete}, FBNSV^{N-}, FBNSV^{U4-}, FBNSV^{C-} and FBNSV^{C-,U4-}) or four (FBNSV^{U2-}) weeks after inoculation. The presence of the segments was controlled by qPCR.

Infection	Infection rates by agro-inoculation*	Symptoms
FBNSV ^{complete}	72/437	Leaf curling, plant stunting
FBNSV ^{N-}	99/212	Leaf curling, plant stunting
	84/178	Leaf curling, plant stunting
FBNSV ^{C-}	38/428	Leaf curling, plant stunting
FBNSV ^{U2-}	1/450**	Attenuated, slight reduction and deformation of the upper leaves, yellowing
FBNSV ^{C-,U4-}	***	Leaf curling, plant stunting

* Plants for which segments other than those voluntarily omitted were absent were not counted as infected.

** Five batches of approximately 90 plants each were inoculated. Of these five batches, only one plant out of a batch of 89 plants had all seven inoculated FBNSV segments (without U2).

*** Plants infected with FBNSV^{C-,U4-} were inoculated with aphids, as explained in the Methods section, so there is no agro-inoculation infection rate.

Table Sup. 6: Statistical analysis of the comparison of the segment relative frequency between complete and incomplete infections

For each type of incomplete infection we first provide the output of a full model, $\text{logit}(\text{freq}) = \text{experiment} * \text{modality} * \text{segment}$ where experiment corresponds to replicate experiments and modality to the incomplete vs complete infection treatment. After the full tests we provide the output of per segment comparisons of the modality to identify segments whose relative frequency statistically significantly differed between incomplete and complete infections. Statistical analysis was assessed through ANOVA and Tukey HSD tests on Logit-transformed data and using JMP. The p-value indicating significant differences after FDR correction (≤ 0.05) are in orange.

FBNSV^{C-}

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	2	2	0.31115	0.3163	0.7291
modality	1	1	0.17480	0.3553	0.5516
experiment*modality	2	2	0.14031	0.1426	0.8672
segment	6	6	172.38127	58.4059	<.0001*
experiment*segment	12	12	6.01695	1.0193	0.4312
modality*segment	6	6	29.87722	10.1229	<.0001*
experiment*modality*segment	12	12	3.62826	0.6147	0.8293

Segment M

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	2	2	0.4798698	0.9438	0.3981
modality	1	1	3.8492036	15.1419	0.0004*
experiment*modality	2	2	0.2167262	0.4263	0.6560

Segment N

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	2	2	0.165719	0.0413	0.9596
modality	1	1	12.684002	6.3260	0.0162*
experiment*modality	2	2	2.181687	0.5440	0.5849

Segment R

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	2	2	3.0767052	6.6148	0.0034*
modality	1	1	2.6057760	11.2046	0.0018*
experiment*modality	2	2	0.2601319	0.5593	0.5763

Segment S

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	2	2	1.5888855	4.5191	0.0174*
modality	1	1	0.0967478	0.5503	0.4627
experiment*modality	2	2	0.4804290	1.3664	0.2673

Segment U1

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	2	2	0.03038378	0.0951	0.9095
modality	1	1	0.83535618	5.2267	0.0279*
experiment*modality	2	2	0.20056869	0.6275	0.5394

Segment U2

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	2	2	0.6854386	1.2295	0.3038
modality	1	1	9.9147912	35.5703	<.0001*
experiment*modality	2	2	0.2906345	0.5213	0.5979

Segment U4

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	2	2	0.30110444	0.4466	0.6431
modality	1	1	0.06613483	0.1962	0.6603
experiment*modality	2	2	0.13838453	0.2052	0.8154

FBNSV^{N-}

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.01223	0.0181	0.8930
modality	1	1	2.09559	3.1022	0.0790
experiment*modality	1	1	0.05576	0.0825	0.7740
segment	6	6	259.97040	64.1418	<.0001*
experiment*segment	6	6	9.05036	2.2330	0.0396*
modality*segment	6	6	21.43553	5.2887	<.0001*
experiment*modality*segment	6	6	2.09372	0.5166	0.7958

Segment C

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	3.9235934	2.6896	0.1072
modality	1	1	1.5344290	1.0518	0.3099
experiment*modality	1	1	0.0513552	0.0352	0.8519

Segment M

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	1.4464297	8.8346	0.0045*
modality	1	1	6.2517240	38.1845	<.0001*
experiment*modality	1	1	0.0879190	0.5370	0.4670

Segment R

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	2.7749930	12.2567	0.0010*
modality	1	1	6.9681318	30.7770	<.0001*
experiment*modality	1	1	0.0405139	0.1789	0.6741

Segment S

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.10198915	0.2403	0.6261
modality	1	1	0.80278431	1.8912	0.1751
experiment*modality	1	1	0.28724606	0.6767	0.4146

Segment U1

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.05706047	0.2118	0.6473
modality	1	1	0.59699143	2.2157	0.1428
experiment*modality	1	1	0.56985085	2.1149	0.1520

Segment U2

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.1299422	0.4528	0.5040
modality	1	1	3.3151917	11.5534	0.0013*
experiment*modality	1	1	1.1083874	3.8627	0.0548

Segment U4

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.6285851	0.3310	0.5676
modality	1	1	4.0618733	2.1392	0.1497
experiment*modality	1	1	0.0042089	0.0022	0.9626

FBNSV^{U2}-

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	1.17681	4.1086	0.0431*
Segment	6	6	244.96941	142.5433	<.0001*
modality*Segment	6	6	84.63332	49.2466	<.0001*

Segment C

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	34.028586	50.8932	<.0001*

Segment M

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	13.356713	69.3222	<.0001*

Segment N

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	4.4383009	25.8773	<.0001*

Segment R

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	0.55919231	1.8217	0.1807

Segment S

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	1.0119220	4.4073	0.0388*

Segment U1

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	4.8243137	36.8531	<.0001*

Segment U4

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	27.591095	90.5482	<.0001*

FBNSV^{U4-}

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment		1	0.22558	0.4042	0.5253
modality		1	0.23214	0.4160	0.5193
experiment*modality		1	0.13725	0.2459	0.6202
segment		6	225.54355	67.3599	<.0001*
experiment*segment		6	5.20945	1.5558	0.1589
modality*segment		6	14.39314	4.2986	0.0003*
experiment*modality*segment		6	4.64282	1.3866	0.2188

Segment C

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	1.9991119	1.7352	0.1933
modality	1	1	0.0040883	0.0035	0.9527
experiment*modality	1	1	0.4626718	0.4016	0.5289

Segment M

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.83837269	5.7248	0.0202*
modality	1	1	0.01577239	0.1077	0.7440
experiment*modality	1	1	0.08628634	0.5892	0.4461

Segment N

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	1.404575	1.0812	0.3031
modality	1	1	12.591589	9.6926	0.0030*
experiment*modality	1	1	2.070624	1.5939	0.2122

Segment R

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.65708034	1.4172	0.2391
modality	1	1	0.81989302	1.7683	0.1892
experiment*modality	1	1	0.15412728	0.3324	0.5666

Segment S

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.18399629	0.6560	0.4215
modality	1	1	0.00128306	0.0046	0.9463
experiment*modality	1	1	0.01367163	0.0487	0.8261

Segment U1

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.00678159	0.0313	0.8602
modality	1	1	0.23956748	1.1070	0.2974
experiment*modality	1	1	0.44076898	2.0367	0.1593

Segment U2

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
experiment	1	1	0.3451167	0.9911	0.3239
modality	1	1	0.9530785	2.7371	0.1038
experiment*modality	1	1	1.5519224	4.4569	0.0394*

FBNSV^{C-, U4-} compared to FBNSV^{Complete}

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	0.687234	2.7830	0.0975
segment	5	5	71.515991	57.9221	<.0001*
modality*segment	5	5	20.201044	16.3612	<.0001*

Segment M

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	0.02076434	0.1654	0.6880

Segment N

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	0.86604356	1.7288	0.2015

Segment R

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	3.5409361	16.3765	0.0005*

Segment S

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	2.7638495	9.5773	0.0051*

Segment U1

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	8.7326332	43.3825	<.0001*

Segment U2

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	4.9640515	33.3043	<.0001*

FBNSV^{C-, U4-} compared to FBNSV^{C-}

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality		1	1.409519	2.6313	0.1064
segment		5	85.801331	32.0349	<.0001*
modality*segment		5	9.022462	3.3686	0.0061*

Segment M

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	0.90930423	3.3386	0.0770

Segment N

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	3.5499833	1.7214	0.1988

Segment R

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	0.40657071	1.6786	0.2044

Segment S

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	2.9224280	12.0623	0.0015*

Segment U1

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	0.95503358	6.1743	0.0184*

Segment U2

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
modality	1	1	1.6886610	7.0280	0.0124*

Table Sup. 7: Sequences of the primers used to quantify segment accumulation by qPCR

Targeted segment	Primer forward (5'-3')	Primer reverse (5'-3')
C	GACCCATCAACTGAAGACCTGTTGTCTC	AGCAACTCTCTACATGCAAGGTCTTCCA
M	ACGCTGCGTATCAAGACGACGG	TCCCGGAAGATAGCATGCAAGCA
N	GTGGTTATCTGTATCATAATGATTATGGCT	CTTCATCTTTCTTCAAATACTGAGACT
R	GGAGAACTGAAGAAGAAGAAAGCTAT	AAGCAAAGCATCAACCTCG
S	AGAAAGATGACGTGTCTGGTAG	TCAGATAATTCACCAACTCTCCCG
U1	AGCTATTGATTTGGAAGACAGAGT	CGACAGCCAAGAAGCGT
U2	ATTAGAGTAGATGAATGCAAACTTGTAT	TAAACCTTCTAGACTGCCATCGT
U4	CCAGATTAAGAAGTTACTGTGTCAGGTAATA	ATTCAGCATCCTCGAACGG

III. Influence des interactions entre les différents composants viraux dans la mise en place de la formule génomique

A. Objectif

Par le système d'agro-infiltrations de feuilles, nous déterminons que l'accumulation de chaque segment génomique à l'échelle locale, en paire avec le R, ne correspond pas à l'accumulation observée lors d'une infection systémique quand tous les segments progressent ensemble. Nous en concluons que la capacité intrinsèque de chaque segment à s'accumuler dans la plante ne permet pas d'expliquer la formule génomique. Ceci implique que la mise en place de la formule nécessite, soit des interactions entre les différents segments génomiques, soit plusieurs cycles d'infection et l'évolution de la composition du groupe de segment au fil de ces cycles. Ainsi, afin de tester la première possibilité, nous avons co-infiltrés les huit segments génomiques dans les mêmes feuilles. Comme expliqué dans la discussion de l'article, nous supposons que dans ce système d'agro-infiltrations les segments génomiques ne sont pas systématiquement, voir sont rarement, infiltrés dans la même cellule et que les échanges de matériel viral (ADN, ARN, protéines) entre cellules sont inexistantes. Il est donc peu probable que les huit segments génomiques puissent interagir dans ce système. Nous pensons que cette expérience ne nous permet pas de conclure sur l'implication des interactions au sein du groupe à l'échelle locale dans l'instauration de la formule génomique.

Afin de déterminer si des interactions entre les segments génomiques suffisent à produire la formule localement, nous avons tenté d'utiliser un système de transfection de protoplastes. Cette approche permettrait d'étudier l'accumulation des segments à l'échelle cellulaire, puisque aucune connexion inter-cellulaire n'existe entre les différents protoplastes en suspension. L'objectif était de transférer des protoplastes avec un mélange équimolaire des huit segments du FBNSV, puis de suivre l'évolution des fréquences relatives des segments au cours du temps. Des fréquences évoluant vers la formule génomique auraient indiqué que celle-ci résulte simplement de la capacité de chaque segment à s'accumuler en interaction avec les autres segments qui composent le groupe, et ne nécessite donc pas de cycles d'infection successifs. Autrement, un pattern de fréquences similaire à celui des infiltrations de feuilles par paires de segments aurait indiqué que la formule génomique ne dépend pas d'interactions directes entre les segments et ne se met en place qu'au fil des cycles d'infection successifs, par exemple par une sélection agissant au niveau du groupe. Enfin, un pattern différent des deux précédents aurait pu

être observé si des interactions au sein du groupe impactent l'accumulation des segments, mais ne suffisent pas à reproduire la formule génomique. Ceci impliquerait que la formule résulte à la fois de la capacité intrinsèque de chaque segment à s'accumuler, d'interactions au sein du groupe et d'une sélection de groupe au cours des cycles d'infection.

B. Méthode

Cette stratégie présentait des prérequis méthodologiques qui n'ont pas pu être remplis en totalité.

1. Optimisation d'un protocole de préparation, transfection et culture de protoplastes de fèves

V. faba est une plante relativement facile à protoplaster et quelques tests sur le matériel végétal à utiliser et les conditions de digestion de la paroi cellulaire ont permis d'établir rapidement un protocole de protoplastation. La productivité des protocoles testés a été estimée par observation et comptage des protoplastes au microscope optique. Plusieurs conditions de cultures ont ensuite été testées afin de déterminer les meilleures conditions de conservation de ces protoplastes.

Dans un second temps, plusieurs paramètres de transfection ont été testés pour établir un protocole le plus optimal possible. Afin d'évaluer l'efficacité de différents protocoles, les protoplastes ont été transfectés avec des plasmides permettant l'expression des gènes rapporteurs *GFP* ou *RFP*. La fluorescence émise par la GFP ou la RFP a permis d'estimer les taux de transfection par simples comptages au microscope à épifluorescence. Les différents paramètres testés (concentrations en protoplastes et en plasmides, agitation, etc.) ont permis d'établir un protocole permettant un taux de transfection de protoplastes d'environ 15 %. Le protocole de protoplastation, transfection et culture des protoplastes est présenté dans l'**Annexe 1**.

2. Transfection du génome viral

Une fois le protocole de transfection mis au point, nous avons tenté de transfecter les segments du FBNSV sous différentes formes :

- un mélange de huit plasmides (pLitmus) contenant chacun la séquence d'un segment du FBNSV sous forme de monomères ou de dimères (ces derniers permettant en théorie la re-circularisation et la réplication avec l'intervention de la protéine Rep) ;

- ou directement les séquences des huit segments linéaires (possiblement recircularisées par la cellule hôte, ce qui générerait des intermédiaires de réplication double-brin circulaires, pouvant à la fois servir de matrice pour la transcription et le traduction de Rep, puis de matrice pour la RCR).

L'ADN total de chaque réaction a été extrait à différents temps après la transfection (0, 24, 42, 66 et 72 heures) et la concentration de chaque segment génomique a été estimée par qPCR. Malheureusement, nous n'avons pas détecté de réplication virale dans nos protoplastes.

C. Perspectives

Par manque de temps, cette approche a été mise de côté. Néanmoins, certains paramètres pourraient encore être testés, notamment le type de matériel viral transfecté. Une possibilité serait d'utiliser des particules virales purifiées à partir de plantes infectées. L'unique étude présentant des transfections de protoplastes avec un nanovirus a été menée par (Franz et al., 1999) en utilisant des particules virales de FBNSV. Le problème majeur est qu'il est difficile d'obtenir des particules virales de FBNSV purifiées et suffisamment concentrées. D'une part, le FBNSV étant restreint aux cellules du phloème, de grandes quantités de matériel végétal sont nécessaires pour obtenir une concentration suffisante de virus purifié. D'autre part, le FBNSV infecte des légumineuses, et nous travaillons notamment avec *V. faba* qui produit des métabolites secondaires en grande quantité, ce qui rend difficile la purification. Nous pourrions envisager de tester d'autres plantes hôtes du FBNSV produisant moins de métabolites secondaires. L'inconvénient de la transfection de particules virales issues de purifications est qu'elle ne permet pas de contrôler directement les fréquences relatives des segments génomiques dans l'*inoculum*. Il faudrait alors considérer les fréquences de départ, correspondant à la formule génomique dans la plante hôte, pour suivre leur évolution dans les protoplastes.

Un autre problème conceptuel de cette approche est qu'elle implique une hypothèse forte loin d'être vérifiée. A cause de l'absence de connexion inter-protoplastes, seule la situation où tous les segments seraient transfectés ensemble dans une même cellule pourrait fonctionner. Il faudrait alors que le groupe de segments se comporte comme en situation naturelle. Or, dans la plante hôte, les segments s'accumulent dans des cellules distinctes. Ainsi, nous ne savons pas si ces fonctionnements uni- et multi-cellulaire peuvent affecter la formule génomique.

IV. Adaptabilité de la formule

Les modifications de formules génomiques observées dans l'expérience d'infections incomplètes supportent l'hypothèse de sélection du ou des groupe(s) de segments pour lesquels les fréquences relatives sont les plus optimales pour le fonctionnement du virus. Néanmoins, ces résultats ne permettent pas de valider directement cette hypothèse. Une manière directe de démontrer le rôle adaptatif de la formule génomique serait d'utiliser le système de transfection de protoplastes décrit précédemment. L'objectif serait de transférer les segments du FBNSV à différentes fréquences relatives dans des protoplastes de fève, puis de déterminer si l'accumulation virale est supérieure dans des protoplastes transfectés avec les fréquences se rapprochant de la formule génomique spécifique de la fève. Par exemple, nous pourrions transférer des protoplastes de *V. faba* avec des particules virales purifiées à partir de *V. faba* ou de *M. truncatula*, puis déterminer si le virus se réplique plus dans les protoplastes transfectés avec la purification issue de *V. faba* (et donc avec un inoculum correspondant à la formule génomique dans *V. faba*). Dans ce dernier cas, les résultats indiqueraient que la formule génomique est adaptative.

La mise en place du système de transfection n'ayant pas abouti, d'autres approches indirectes ont été imaginées afin de tester la fonctionnalité adaptative de la formule génomique. Ces approches visaient à perturber le fonctionnement du groupe de différentes manières et d'en déterminer l'impact sur la formule génomique. La première stratégie consistait à produire de manière constitutive *in planta* une des protéines du FBNSV. La seconde cherchait à modifier l'impact possible des mécanismes de *silencing* dirigés contre certains segments du FBNSV.

A. Expression de protéines du FBNSV *in planta*

Nous supposons que la formule génomique résulte de la sélection des fréquences relatives les plus optimales pour le fonctionnement du virus. Si les besoins fonctionnels du virus changent, plus particulièrement si la fonction d'un de ses segments est remplie par l'hôte, alors ce segment devient inutile pour le groupe, ou tout au moins son expression doit être ajustée. Il serait alors probablement perdu au fil des cycles d'infection puisque sa réplication et son encapsidation représente un coût pour le virus. Alternativement, sa fréquence dans la formule pourrait être réduite afin d'ajuster l'expression du gène

correspondant à un niveau acceptable. Ceci ne s'applique pas aux segments N et U4 qui, comme discuté dans l'article, ne semblent pas avoir de rôle dans l'infection de *V. faba* et ont une capacité d'accumulation supérieure aux autres segments. Ainsi, l'objectif de cette expérience était de produire des plantes exprimant une des protéines du FBNSV de manière constitutive, de les infecter avec les huit segments du FBNSV, puis d'étudier l'impact de cette expression sur le segment correspondant ainsi que sur les autres segments.

Pour créer ce système d'expression *in planta*, l'utilisation de *V. faba* transgéniques ne semblait pas possible car cette plante n'étant pas une espèce modèle, nous ne disposons pas de méthode de transformation stable. Dans un premier temps, nous avons pensé utiliser *Arabidopsis thaliana* comme plante hôte, puisque ces plantes sont facilement et rapidement transformables (Bechtold and Pelletier, 1998; Clough and Bent, 1998). De plus, *A. thaliana* peut-être infectée par le nanovirus FBNYV (Vega-Arreguín et al., 2007). Malheureusement, après plusieurs essais, nous ne sommes pas parvenus à infecter *A. thaliana* avec le FBNSV. Notre seconde stratégie était d'utiliser un autre virus comme vecteur d'expression de protéines du FBNSV. Pour cela, il fallait identifier un virus candidat, insérer la séquence d'un gène du FBNSV dans le génome de ce vecteur et tester la production de la protéine dans la plante, avant d'inoculer le FBNSV.

1. Choix du virus vecteur d'expression

Le virus vecteur d'expression devait remplir plusieurs conditions :

- (i) pouvoir infecter *V. faba* mais ne pas causer de symptômes sévères,
- (ii) être disponible sous forme de clone infectieux pouvant être facilement modifié, et
- (iii) pouvoir être co-infecté avec le FBNSV et interférer le moins possible avec celui-ci.

Après des recherches bibliographiques, nous avons choisi de travailler avec des potyvirus. Les membres de cette famille infectent un large spectre d'hôtes, y compris des *Fabaceae*. Il s'agit de virus à (+)ssRNA qui se répliquent dans le cytoplasme de leur cellule hôte. Ainsi, leur réplication pourrait ne pas interférer avec celle du FBNSV qui possède un génome à ADN et se réplique dans le noyau. Le principal point fort des potyvirus est que plusieurs d'entre eux ont été utilisés pour l'expression de protéines *in planta* dans le cadre de plusieurs études indépendantes. Cependant, ils présentent aussi deux désavantages majeurs. D'abord, ils infectent tous les tissus de la plante et l'expression constitutive des protéines de nanovirus dans certains tissus pourrait être toxique pour la plante. Nous avons notamment observé que l'agro-infiltration du segment R dans des feuilles de *N. benthamiana* entraîne leur nécrose. D'autre part, l'expression des gènes des potyvirus passe par la production d'une polyprotéine ensuite clivée par des

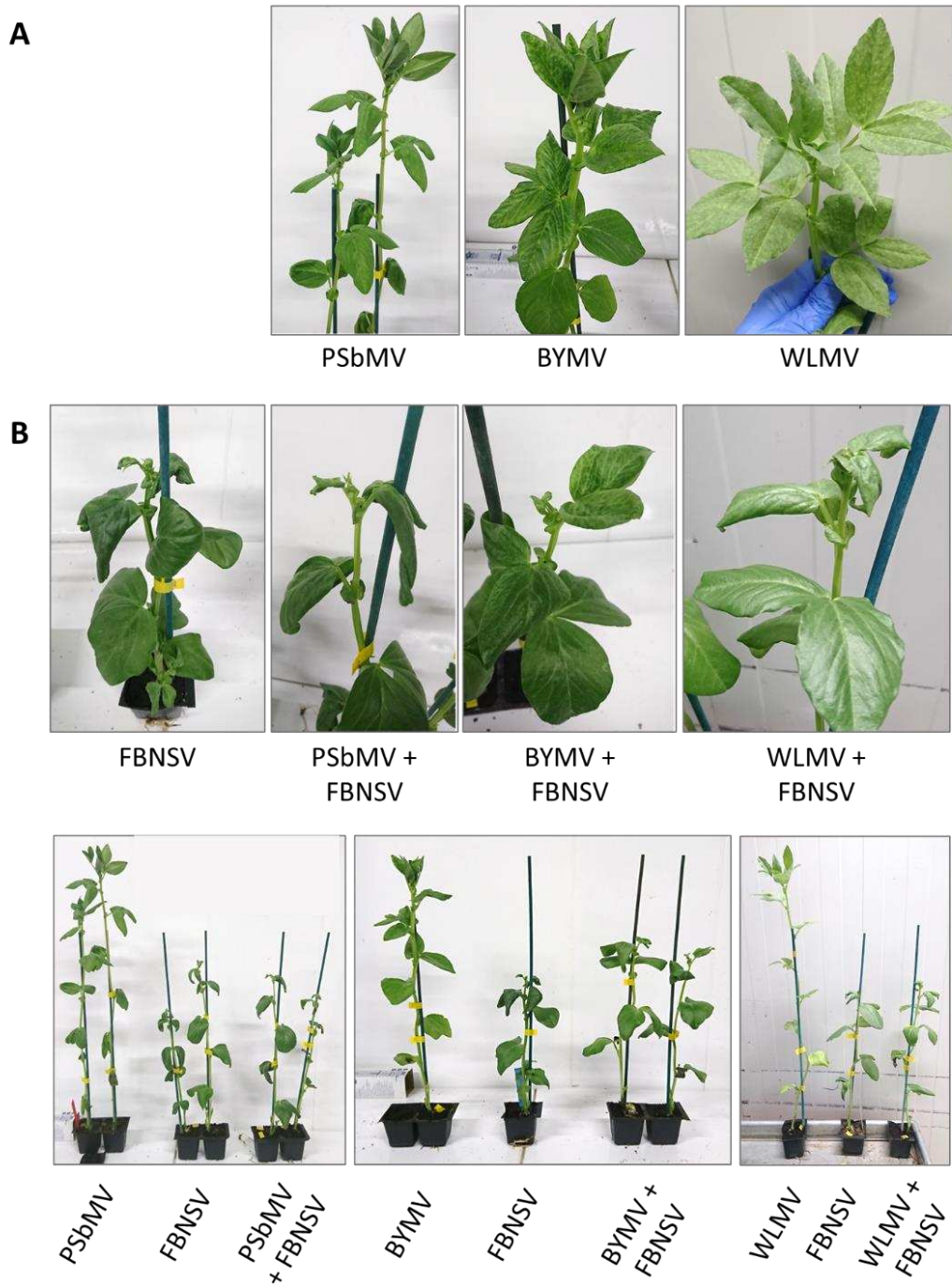


Figure 3 : Phénotypes de *V. faba* infectées par un des potyvirus PSbMV, BYMV ou WLMV, par le FBNSV ou co-infectées par un potyvirus et le FBNSV. Les photos ont été prises trois semaines après l'inoculation. Le PSbMV est asymptomatique, alors que le BYMV et le WLMV provoquent des décolorations en mosaïques sur des feuilles. Le FBNSV induit un nanisme et un enroulement des feuilles à l'apex des étages foliaires supérieurs. Les plantes co-infectées présentent une combinaison des symptômes des deux virus, sans synergie ou antagonisme apparent.

protéinases virales qui reconnaissent des motifs spécifiques d'acides aminés. Cette stratégie d'expression est différente de celle des nanovirus qui expriment leurs gènes avec des transcrits indépendants, et pourrait influencer sur les modifications post-traductionnelles des protéines produites. De plus, après le clivage, quelques acides aminés constituant le motif de reconnaissance par la protéinase restent à l'extrémité de la protéine libérée. Pour ces raisons, la conformation et la fonctionnalité des protéines du FBNSV pourraient être altérées par le système d'expression du potyvirus.

Quatre potyvirus infectant les légumineuses et pour lesquels des clones infectieux étaient disponibles ont été sélectionnés : le pea seed-born mosaic Virus (PSbMV ; pathotype P4 ; (Johansen et al., 2001, 1996)), le bean common mosaic virus (BCMV ; (Naderpour and Johansen, 2011)), le bean yellow mosaic virus (BYMV ; (Guyatt et al., 1996)) et le white lupin mosaic virus (WLMV ; (Bruun-Rasmussen et al., 2007)). Des clones infectieux de ces quatre virus nous ont été fournis par E. Johansen (Université de Copenhague). Nous avons vérifié l'infectivité de ces virus sur *V. faba*, ce qui nous a permis d'éliminer le BCMV qui s'est avéré ne pas infecter la fève. Le BYMV et le WLMV provoquent des symptômes de mosaïque, avec un léger épaissement des feuilles pour la BYMV. Le PSbMV est asymptomatique (**Figure 3A**).

2. Co-infections avec le FBNSV

Chaque potyvirus a été inoculé avec le FBNSV afin de s'assurer que la co-infection soit possible et n'affecte pas l'infection du nanovirus. Ces expérimentations étaient uniquement destinées au choix du virus vecteur, et seulement quelques plantes infectées par le FBNSV seul ont été produites pour comparaison. Les résultats nous informent donc sur d'éventuelles modifications drastiques de l'infection du FBNSV par la présence du potyvirus, mais pas sur de plus petites variations. Cependant, ils sont suffisants pour répondre à notre problématique.

Les potyvirus ont été inoculés mécaniquement avec un extrait brut de tissus infectés broyés dans un tampon de sodium phosphate (0,04 M) additionné de poudre de carborundum. Le FBNSV a été inoculé par puceron (*A. pisum*) soit immédiatement après le potyvirus, soit sept jours plus tard, dans l'optique de produire la protéine nanovirale avant l'inoculation du FBNSV. Dans les deux conditions, la quasi-totalité des plantes co-inoculées étaient infectées par les deux virus, indiquant que l'infection par l'un n'exclut pas l'autre. Les symptômes des plantes co-infectées correspondaient à l'addition des symptômes des deux maladies (**Figure 3B**), sans aucune synergie ou aucun antagonisme apparent. Les charges totales de FBNSV en présence ou non du potyvirus étaient assez similaires (**Figure 4**). Au niveau de la formule génomique, la co-infection n'induisait pas de changement qualitatif du pattern d'accumulation des segments

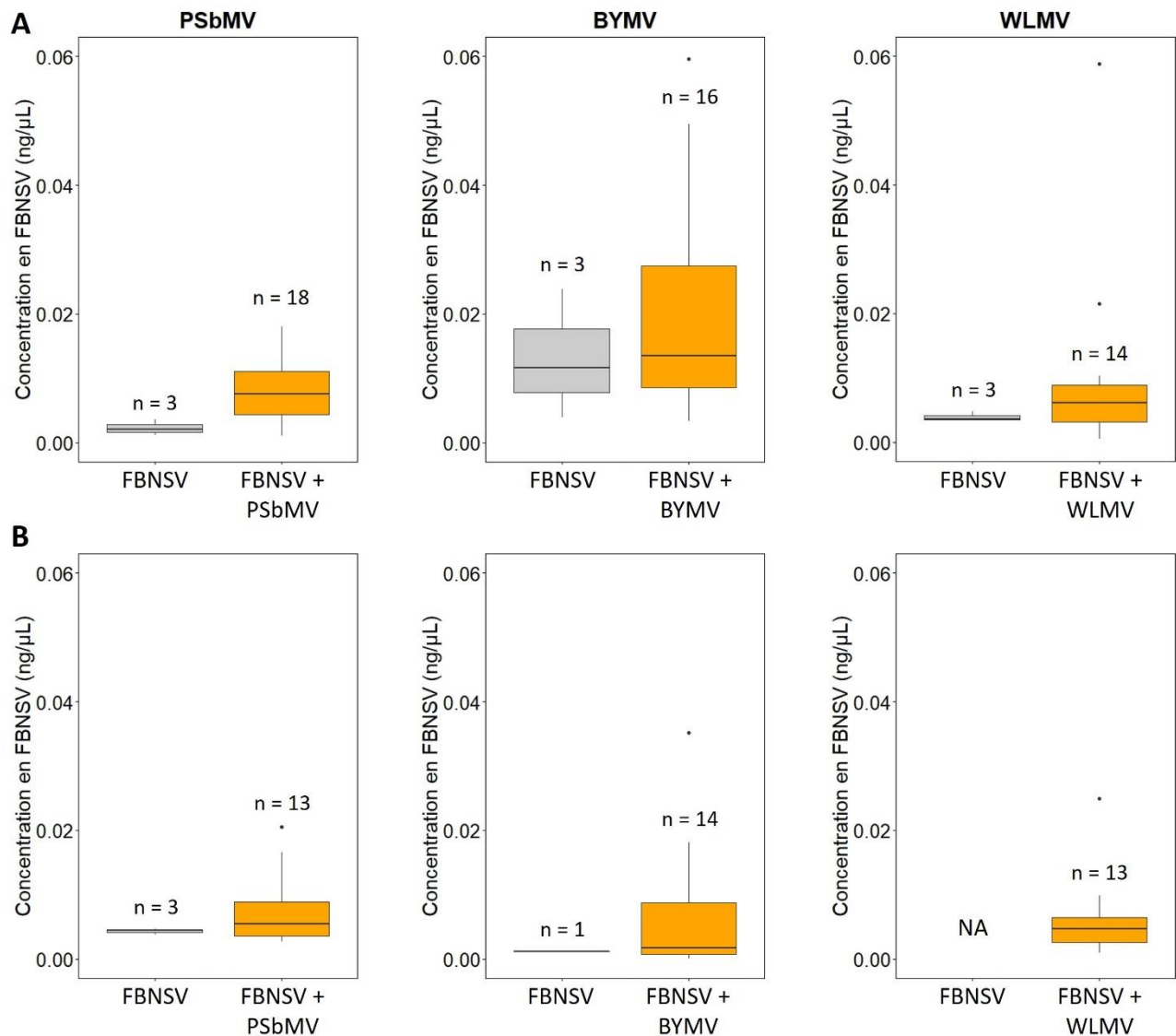


Figure 4 : Concentration en FBNSV dans *V. faba* lorsqu'il est inoculé seul (gris) ou avec un potyvirus (orange). Le nom du potyvirus est indiqué au-dessus de chaque graphique. Le FBNSV a été inoculé en même temps que (A) ou sept jours après (B) le FBNSV. La charge totale de FBNSV a été estimée par qPCR trois semaines après son inoculation. La barre centrale représente la médiane et les bords du rectangle représentent le premier et le troisième quartiles. Les barres extérieures délimitent 1,5 fois la distance entre les premier et troisième quartiles de la distribution. Les points représentent les valeurs extrêmes. « n = » indique la taille d'échantillon.

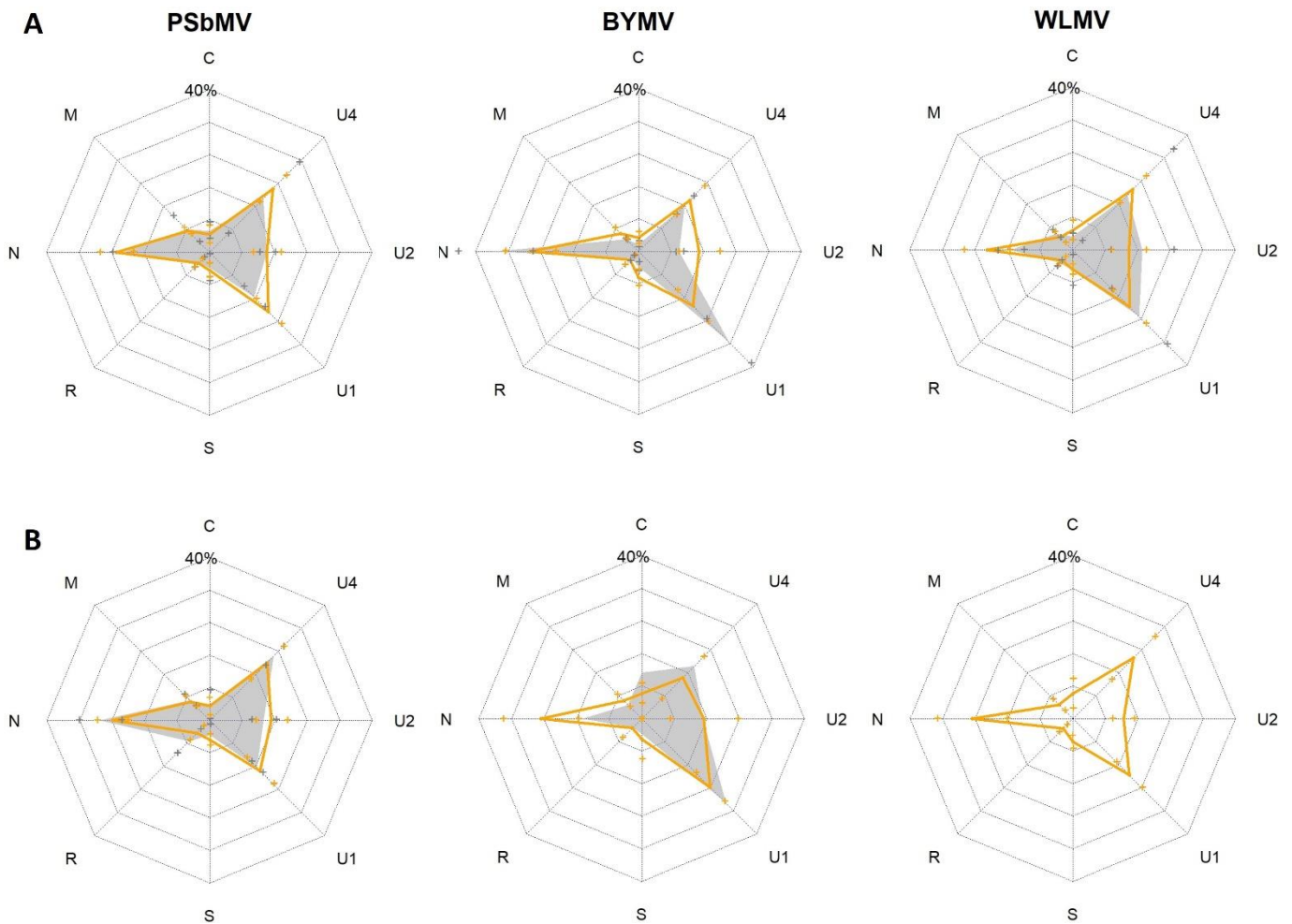


Figure 5 : Comparaison des formules génomiques du FBNSV seul (gris) ou en co-infection avec un potyvirus (orange) dans *V. faba*. Le nom de chaque potyvirus est indiqué au-dessus du radarplot. Le FBNSV a été inoculé en même temps que **(A)** ou sept jours après **(B)** les autres virus. L'accumulation des segments génomiques a été estimée par qPCR et la fréquence de chaque segment dans la population virale a été déterminée. La médiane des fréquences de chaque segment est indiquée sur l'axe correspondant. Les écart-types sont représentés par les croix. Les formules génomiques ont été estimées sur les mêmes plantes que celles de la **Figure 4**.

génomiques (**Figure 5**). Ces résultats confirment qu'il n'y a pas de synergie ni d'antagonisme entre les infections du FBNSV et chacun des trois potyvirus. Puisqu'aucun des trois potyvirus ne semblait affecter drastiquement l'infection par le FBNSV, nous avons choisi de travailler avec le PSbMV car il n'engendre pas de symptômes sur *V. faba*.

3. Stratégie de clonage de séquences codantes du FBNSV dans le clone infectieux du PSbMV

Comme mentionné précédemment, l'expression des gènes des potyvirus passe par la synthèse d'une grande polyprotéine, qui est ensuite clivée pour libérer les différentes protéines virales. Le clivage protéolytique est effectué par les protéinases du virus qui reconnaissent des séquences d'acides aminés spécifiques. Nous avons choisi d'insérer les séquences du FBNSV entre les séquences des gènes codant pour les protéines Nib et CP (**Figure 6**). Notre stratégie était de créer un site d'insertion dans la séquence nucléotidique en substituant deux nucléotides dans le motif de reconnaissance par la protéinase et en aval du site de clivage (**1**). Ces substitutions synonymes modifient la séquence nucléotidique pour créer un motif reconnu par l'enzyme de restriction NaeI, mais pas la séquence d'acides aminés reconnue par la protéinase du PSbMV. Une digestion par NaeI à ce site aurait ensuite permis l'ouverture du plasmide (**2**) et l'insertion de la séquence du FBNSV flanquée à l'extrémité C-terminale par la séquence codante d'un site de clivage protéolytique supplémentaire (**3**), de sorte que la protéine du nanovirus soit clivée aux deux extrémités. Malheureusement, le clone du PSbM s'est avéré très instable et malgré l'utilisation de diverses techniques, nous ne sommes pas parvenus à créer le site de restriction enzymatique. Nous avons également essayé des stratégies similaires sur les clones infectieux du BYMV et du WLM. Nous avons rencontré le même problème d'instabilité des plasmides.

4. Discussion et perspectives

Cette approche n'a donc pas abouti. Néanmoins, la production de plantes exprimant des protéines du FBNSV fournirait un outil utile pour répondre à différentes problématiques concernant le fonctionnement et la transmission du FBNSV.

L'utilisation d'un virus comme vecteur a l'avantage de fournir rapidement des plantes produisant une protéine d'intérêt. De plus, une fois la construction « de base » disponible, il est possible de générer facilement de nombreux clones permettant l'expression de différents gènes par simples clonages. Par exemple, il serait possible de tester toutes les protéines de nanovirus en changeant uniquement la séquence nanovirale contenue dans le vecteur. Néanmoins, cette approche est dépendante de paramètres liés à la biologie du virus vecteur (spécificité tissulaire, taille du génome, modifications post-

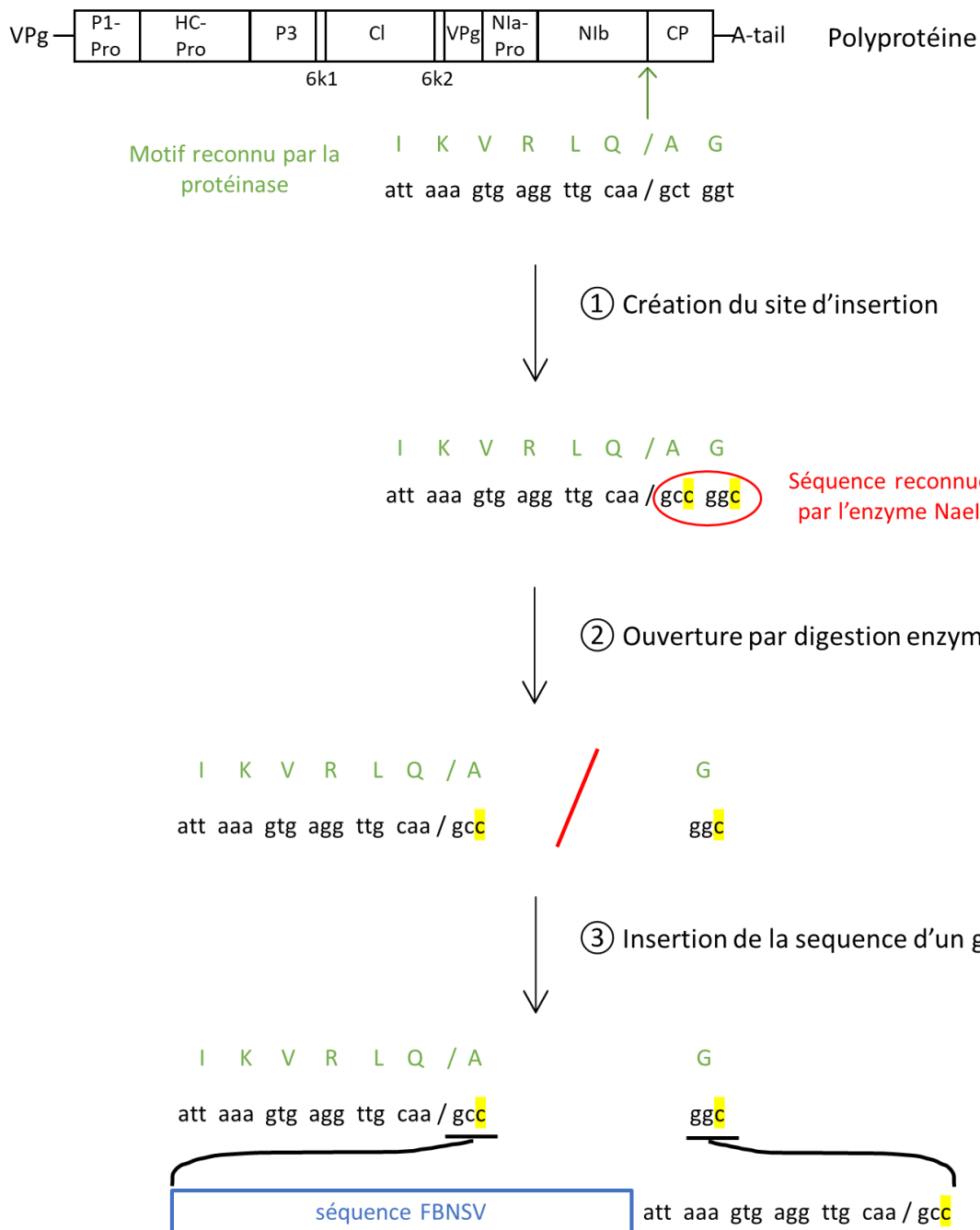


Figure 6 : Stratégie d'insertion de la séquence codante d'un segment du FBNSV dans le génome du PSbMV. (1) Création d'un site de clivage par l'enzyme de restriction NaeI (cercle rouge) par substitution de deux nucléotides (jaunes) dans le motif de reconnaissance par la protéinase (vert) et en aval du site de clivage (représenté par le « / »). (2) Ouverture du plasmide par digestion de l'enzyme NaeI et (3) insertion de la séquence du FBNSV (rectangle bleu) flanquée à l'extrémité C-terminale par la séquence codante d'un site de clivage protéolytique supplémentaire.

traductionnelles, insertion de séquences non désirées, *etc.*) qui génèrent des incertitudes à diverses étapes de la stratégie.

Une autre possibilité serait d'utiliser une autre plante hôte. *M. truncatula* est une espèce modèle pour laquelle il existe des protocoles de transformation stable. Le FBNSV peut l'infecter et sa formule génomique dans cet hôte est connue (Gallet et al., 2022; Sicard et al., 2013). Cette approche nécessite un temps de génération des plantes exprimant les protéines du FBNSV plus long que la stratégie du vecteur viral, mais elle présente d'autres avantages. Elle permet par exemple de choisir le promoteur qui contrôlera l'expression du transgène. Pour répondre à notre problématique, il serait plutôt judicieux de conserver les promoteurs endogènes des segments du FBNSV, particulièrement afin de conserver la spécificité tissulaire de l'expression puisque la plupart des promoteurs de *Nanoviridae* sont principalement actifs dans les cellules associées au phloème (Dugdale et al., 2000, 1998; Shirasawa-Seo et al., 2005).

B. Formule et *silencing* antiviral

1. Principe

Par cette approche, nous voulions regarder ce qu'il se passe au sein du système viral FBNSV lorsqu'un des segments a un effet négatif sur un autre segment. Pour cela, nous avons tenté d'exploiter le mécanisme de *gene silencing* utilisé par la plante pour lutter contre l'infection virale. Le *silencing* antiviral génère des petits ARN interférents (siRNA) à partir d'ARN double brin provenant du virus ciblé. Ces siRNA sont intégrés à des complexes dans lesquels ils servent de guide pour la reconnaissance des séquences virales. Ces complexes répriment l'expression des gènes viraux en agissant au niveau transcriptionnel ou post-transcriptionnel. L'objectif était de modifier un des segments du FBNSV de façon à ce qu'il génère plus de siRNA dirigés contre un autre segment, puis de déterminer l'impact que cela a sur l'accumulation ces deux segments.

2. Méthode

Chez des plantes *V. faba* infectées par le FBNSV, l'accumulation de siRNA ciblant chaque segment génomique est proportionnelle aux fréquences relatives observées pour le segment d'ADN auquel ils correspondent (Pooggin, résultats non publiés). Ainsi, le segment U4, qui a une des fréquences relatives les plus élevées dans cet hôte, génère plus de siRNA que les segments M et U2. Notre stratégie était de remplacer une partie de la séquence 3'-UTR du segment U4 par des séquences courtes (environ 30 nucléotides) provenant des segments M ou U2, afin d'augmenter la quantité de siRNA dirigé contre ces

Tableau 1 : Segments et régions ciblées par les constructions contenant le segment U4 modifié

La construction C3 contient une sous unité de la v-ATPase qui est une enzyme exprimée dans le tube intestinal des insectes et nécessaire à leur survie. La SCAMP est une protéine suspectée de jouer un rôle dans la transmission du FBNSV (Jérémy Di Mattia et Jean-Louis Zeddani, résultats non-publiés).

Construction	Cible	Région ciblée
MB1	Segment M	Région codante
MB2		Région promotrice
MB3		3'-UTR
MB4	Segment U2	Région codante
MB5		Région promotrice
MB6		3'-UTR
C3	v-ATPase	Région codante
C5	SCAMP	Région codante

deux segments par rapport à une infection normale. La région 3'-UTR du segment U4 a été choisie pour le remplacement de séquence car elle n'est pas transcrite et ne contient pas de séquences régulatrices connues. Sa modification est donc moins susceptible d'affecter la réplication du segment U4 et l'expression du gène pour lequel il code.

Les zones génératrices de siRNA n'étant pas uniformément réparties sur les segments, trois constructions ciblant la séquence codante, la région promotrice ou le 3'-UTR ont été créés pour chacun des segments ciblés, M et U2 (**Tableau 1**). Deux constructions supplémentaires, contenant chacune un gène de puceron et ne devant ainsi pas affecter la production de siRNA dirigés contre le FBNSV, ont été synthétisées comme contrôles. Chacune de ces constructions a été utilisée pour transformer une colonie d'*A. tumefaciens* (souche COR308) afin d'être agro-inoculée.

Des plantules de *V. faba* âgées de neuf jours ont été agro-inoculées avec les segments C, M, N, R, S, U1, U2 du FBNSV et l'une des six constructions de U4 modifiées selon le protocole décrit par (Sicard et al., 2013). Des plantes témoins inoculées avec U4 non modifié ou modifié avec une séquence d'un gène de puceron ont été cultivées en parallèle. Trois semaines après l'inoculation, l'ADN total a été extrait des plantes symptomatiques en suivant le protocole décrit par (Di Mattia et al., 2020a). L'accumulation de chaque segment du FBNSV a été déterminée par qPCR comme décrit dans l'article. Les formules génomiques pour les infections avec les différentes constructions (FBNSV^{MB1} à FBNSV^{MB4}, FBNSV^{C3} FBNSV^{C5} et FBNSV^{U4}) ont été déterminées en exprimant la concentration de chaque segment génomique relativement à la concentration virale totale dans chaque plante.

Afin de vérifier que des siRNA soient bien produits à partir des séquences incorporées dans la séquence de U4, l'ARN total de plantes infectées a été extrait selon le protocole adapté de (Morante-Carriel et al., 2014) et les siRNA ont été séquencés (Illumina).

3. Résultats et discussion

Le temps d'apparition et la sévérité des symptômes provoqués par les différentes infections ne différaient pas de ceux d'une infection classique (résultats non montrés). La présence d'un insert dans la séquence de U4 a tendance à diminuer la fréquence relative de U4 et augmente celle de N par rapport au FBNSV sauvage (**Figure 7** ; **Tableau 2**). Nous avons également constaté une augmentation parfois significative de la fréquence du segment S. Ces variations (statistiquement significatives ou non) sont obtenues avec tous les segments U4 modifiés, y compris ceux avec des séquences non virales. Ainsi, les modifications de formules génomiques observées semblent provenir de la modification de la séquence du segment U4 plutôt que spécifiquement du *silencing* antiviral sur les segments M et U2 ciblés. Nous

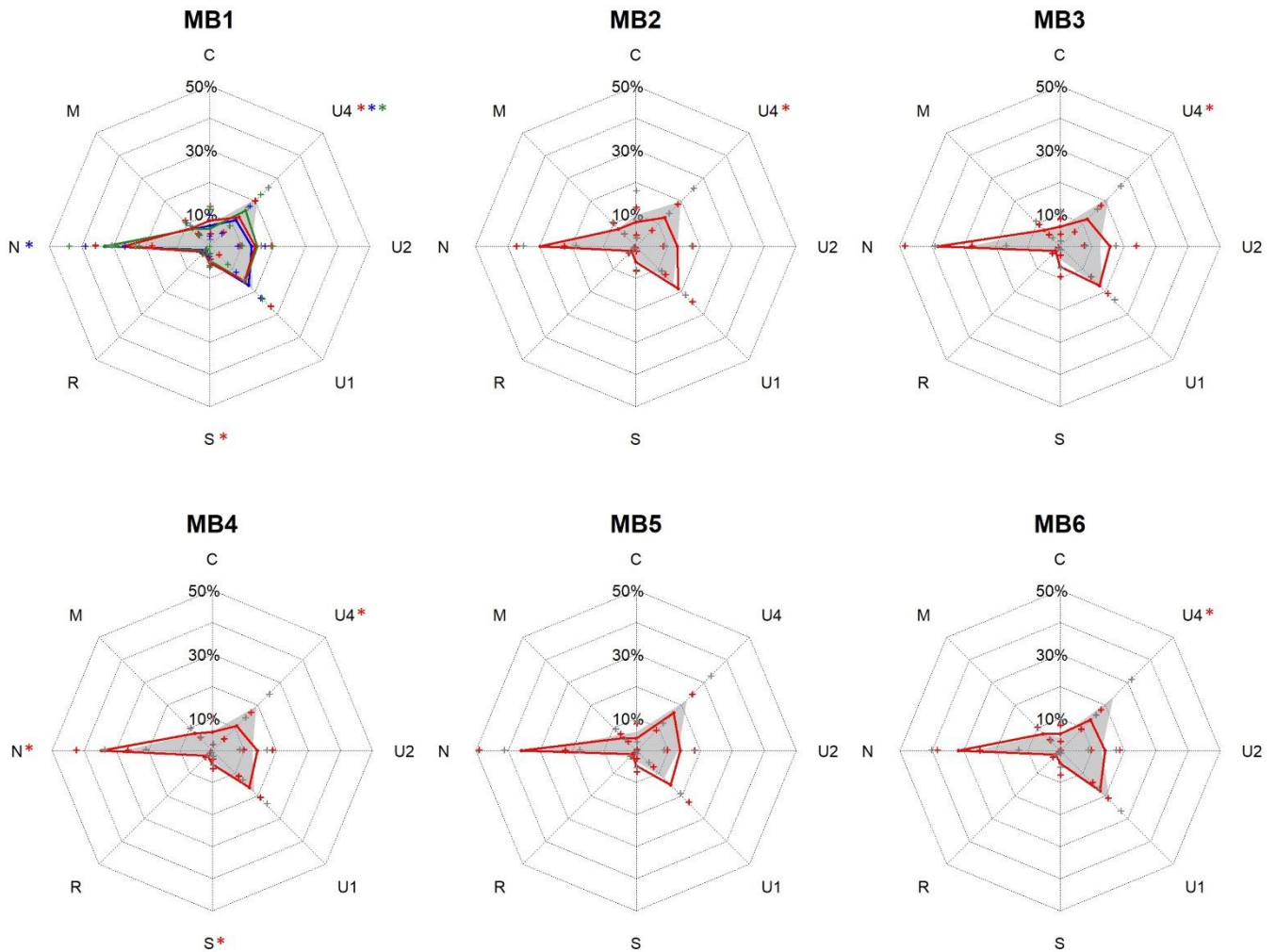


Figure 7 : Comparaison des formules génomiques du FBNSV^{U4} (gris), FBNSV^{MB} (rouge), FBNSV^{C3} (bleu) et FBNSV^{C5} (vert) dans *V. faba*. L'accumulation des segments génomiques a été estimée par qPCR et la fréquence de chaque segment dans la population virale a été déterminée. La médiane des fréquences de chaque segment est indiquée sur l'axe correspondant. Les écart-types sont représentés par les croix. Les astérisques associées au nom des segments indiquent lorsque les différences de fréquences entre FBNSV^{U4} et FBNSV^{MB} (rouge) ou FBNSV^{C3} (bleu) ou FBNSV^{C5} (vert) sont significatives. Les différences entre les segments ont été évaluées par une ANOVA ($p \leq 0,05$). Les résultats des test statistiques sont présentés dans le **Tableau 2**.

supposons que la modification de U4 affecte sa capacité à se répliquer dans la plante et ainsi sa fréquence relativement aux autres segments. Par ailleurs, dans la discussion de l'article, nous proposons que N et U4 soient plus compétitifs que les autres segments pour la réplication dans *V. faba*. Nous pouvons supposer qu'il y ait une compétition entre ces deux segments pour la réplication. Ainsi, les segments U4 modifiés étant moins bien répliqués que le sauvage, ils sont moins compétitifs pour la réplication face au segment N. Le segment N se répliquerait alors plus que dans des plantes FBNSV^{U4}, d'où la tendance à une augmentation de la fréquence du segment N dans les plantes infectées avec des constructions de U4 modifié (**Figure 7**). En revanche, nous n'expliquons par l'augmentation du segment S.

Enfin, le séquençage des ARN montre que des siRNA sont bien produits à partir des séquences introduites dans U4. Des analyses sont en cours afin de vérifier que ces siRNA permettent le clivage des transcrits des segments ciblés. En attendant, nous ne savons pas si les modifications du segment U4 ont impacté l'expression de M ou U2. Nous pouvons uniquement conclure que le remplacement de la petite fraction de 30 nucléotides dans le segment U4, indépendamment de la séquence qui la remplace, impacte négativement l'accumulation du segment.

Tableau 2 : Analyse statistique des comparaisons des fréquences relatives des segments génomiques entre FBNSV^{U4}, FBNSV^{MB}, FBNSV^{C3} et FBNSV^{C5} dans *V. faba*.

Les comparaisons des fréquences relatives des segments ont été réalisées par ANOVA. « dF » indique le nombre d'observations et la F-value la valeur de la statistique F. Les p-values indiquant des différences statistiquement significatives ($\leq 0,05$) après correction FDR sont indiquées en rouge.

	Segment	C	M	N	R	S	U1	U2	U4
C3	Df	38							
	F-value	1.148	1.155	18.48	0.945	6.624	0.011	0.988	23.87
	p-value	0.291	0.289	0.000115	0.337	0.0141	0.916	0.326	1.9e-05
C5	Df	38							
	F-value	0.199	3.114	3.162	0.355	4.368	0.338	2.455	15.51
	p-value	0.658	0.0857	0.0834	0.555	0.0434	0.565	0.125	0.000339
MB1	Df	32							
	F-value	0.048	2.069	0.776	2.989	7.592	0.001	1.012	8.844
	p-value	0.829	0.16	0.385	0.0935	0.00959	0.975	0.322	0.00555
MB2	Df	47							
	F-value	1.706	2.308	5.188	0.506	0.628	2.182	0.751	14.860
	p-value	0.198	0.135	0.0273	0.481	0.432	0.146	0.391	0.000351
MB3	Df	31							
	F-value	3.086	0.059	6.832	0.492	4.87	2.593	4.198	37.05
	p-value	0.888	0.809	0.0137	0.488	0.0349	0.117	0.049	9.59e-07
MB4	Df	36							
	F-value	0.397	0	17.79	2.453	7.819	0.656	0.264	32.92
	p-value	0.533	0.986	0.000159	0.126	0.00825	0.423	0.264	1.56e-06
MB5	Df	40							
	F-value	0.274	1.595	0.936	0.014	2.71	0.397	0.133	2.402
	p-value	0.604	0.214	0.339	0.908	0.108	0.532	0.717	0.129
MB6	Df	25							
	F-value	0.316	1.433	6.973	3.789	1.972	1.637	0.804	29.71
	p-value	0.579	0.243	0.0141	0.0629	0.173	0.212	0.378	17e-05

Chapitre 3 : Etude de l'impact de l'état physiologique de la plante hôte sur la formule génomique

I. Contexte

La formule génomique est dépendante du génotype de l'hôte puisqu'elle change d'une espèce à une autre. Par ailleurs, la plante peut être sujette à des conditions environnementales qui impactent son métabolisme. Dans ce chapitre, nous avons ainsi cherché à déterminer quel est l'influence de l'état physiologique de l'hôte sur la formule génomique. Pour cela, nous avons déterminé la formule du FBNSV dans de fèves cultivées dans différentes conditions de culture.

Nous avons choisi de rédiger cette partie sous forme d'article bien qu'il ne soit pas sous sa forme définitive. Des expériences complémentaires sont encore en cours. De ce fait, la section classiquement intitulée « Discussion » soit ici intitulée « Discussion and perspectives ».

II. Article : Impact of the physiological state of the host plant on the genome formula of a nanovirus

Introduction

Organisms must adapt their gene expression to their environment. This adaptation can be achieved by modifications in the coding or regulatory sequences. However, adaptive mutation rate can be low and/or selection to fixation can be slow, and this process may sometime fail to allow very rapid adaptation of gene expression (Berg et al., 2004; Tuğrul et al., 2015). Another phenomenon, considered to be a fast coarse response is the regulation of gene expression by copy number variation (CNV) (Elde et al., 2012). The CNV mechanism is reversible and allows rapid adaptation of gene expression to changing environments (Tomanek et al., 2020). It relies on gene amplification and contraction and has been named genomic or genetic accordion (Bayer et al., 2018; Elliott et al., 2013). Adaptive CNV is an ubiquitous mechanism and has been empirically validated in various systems, such as insects (Guillemaud et al., 1999; Raymond et al., 1998), fungi (Todd and Selmecki, 2020), bacteria (Elliott et al., 2013), and large double-stranded DNA monopartite viruses that can easily accommodate large genome size variations (Bayer et al., 2018).

In viruses with a segmented genome, it has been observed that the different genome segments reproducibly accumulate at different relative frequencies in the host (Bashir et al., 2022; Diefenbacher et al., 2018; Guyot et al., 2022; Hu et al., 2016; Mansourpour et al., 2022; Moreau et al., 2020; Sicard et al., 2013; Wu et al., 2017; Yu et al., 2019, Yvon et al., 2023). This phenomenon, known as genome formula (Sicard et al., 2013), is believed to play a role analogous to CNV or genomic accordion (Gallet et al., 2022; Gutiérrez and Zwart, 2018; Sicard et al., 2016, 2013; Zwart and Elena, 2020).

It has been established that the genome formula is determined by the host genotype, since the relative frequency of genome segments changes upon host switching, thus upon environmental changes (Gallet et al., 2022; Moreau et al., 2020; Sicard et al., 2013; Wu et al., 2017). However, the importance of the physiological state of a given host genotype, which is dependent on its environment, in modulating the genome formula has not been studied. Viruses are dependent on their host cell to complete their infection cycle. Changes in the plant's physiology, particularly those associated with biotic or abiotic stresses, could affect viral infection by modifying the interactions between the virus and cellular

components, at least quantitatively, and this could be accommodated through modification of the genome formula.

We used the faba bean necrotic stunt virus (FBNSV; family *Nanoviridae*, genus *Nanovirus*), an aphid transmitted virus infecting vascular tissues of many legume plants, to determine whether physiological changes in the plant can impact on the genome formula. FBNSV is a multipartite virus whose genome is divided into eight genome segments (C, M, N, R, S, U1, U2, U4) each encapsidated individually (Gronenborn, 2004; Lal et al., 2020). We studied the relative frequency of FBNSV segments in faba beans (*Vicia faba*) grown under a variety of culture conditions. We chose abiotic and biotic conditions based on their impact and the responses they generate in the plant. The perception of stress leads to classic plant responses, notably the production of calcium waves, ROS burst and phytohormones production or modulation. However, while these responses are common to different stresses, their trigger, the pathways they involve downstream and the final response of the plant are specific to each stress (Zhang et al., 2022), involving different players who can potentially interfere with viral infection.

Most studied abiotic conditions are water deficit, high salt concentration and temperature. Their effects on plants and on the signaling pathways they involve are extensively documented (Lamers et al., 2020; Zhang et al., 2022). Water deficiency generates hyperosmotic stress, which can lead to plasmolysis and detachment of the plasma membrane from the cell wall. In addition to hyperosmotic stress, high salt concentrations lead to an accumulation of ions that are toxic to the cells. Temperature can influence the physical properties of molecules, thus modifying RNA secondary structures, protein stability, conformation and function, as well as membrane properties (Lamers et al., 2020; Zhang et al., 2022).

Biotic stresses could interact with FBNSV either indirectly via a modification of the plant's physiological state and in particular its defense systems, or directly via competition for cellular resources. The stress generated by an aphid infestation could indirectly interfere with FBNSV, since the perception of aphid punctures activates the plant's defense mechanisms, including a reinforcement of structural barriers through callose deposition (Jaouannet et al., 2014) and occluding the phloem sieve tube through polymerization of P proteins (Lewis et al., 2022) or contraction of the forisome (Medina-Ortega and Walker, 2015; Will et al., 2007), thus blocking sap flow. The co-infection with another virus could also directly or indirectly affect FBNSV infection. *Bean yellow mosaic virus* (family *Potyviridae*, genus *Potyvirus*) is a single-stranded RNA virus which, upon infection of the plant, modifies the cellular environment to its advantage (Zhou et al., 2009). It attenuates the plant's defense mechanisms, notably by suppressing the plant's gene silencing mechanism (Valli et al., 2018). Numerous cases of synergy linked to the suppression

of silencing by potyviruses have been described during co-infections with other viruses (Mascia and Gallitelli, 2016). Infection of the plant by BYMV could therefore generate beneficial modifications for infection by FBNSV. Conversely, the two viruses could compete for certain cellular resources, such as the plasmodesmata they use to move from cell to cell (Revers and García, 2015).

Finally, growing plants in symbiosis with *Rhizobia* has both a biotic stress effect, *via* interaction with the bacteria, and an abiotic effect, *via* nitrogen nutrition. The symbiotic interaction between plant and *Rhizobia* induces a systemic response in the plant that induces changes in defense mechanisms and altered metabolism. These changes have been shown to increase plant tolerance to viral infections (Abdelkhalek et al., 2022; Andreola et al., 2019). On the other hand, by converting atmospheric nitrogen into a form of nitrogen that can be assimilated by the plant, *Rhizobia* provide the plant with nitrogen. Nitrogen is an essential component of the cell, necessary for the completion of numerous metabolic functions and thus indispensable for plant growth and development.

Overall, our results show that unlike host genotypes which have major impacts on the FBNSV genome formula (Gallet et al., 2022; Sicard et al., 2013), the various physiological conditions of the host plant induced by a diversity of stresses appear to have only marginal effects.

Material and methods

FBNSV clone and *Agrobacterium*-mediated inoculation

The FBNSV infectious clone (isolate JKI-2000) was provided by (Grigoras et al., 2009) and is composed of eight binary plasmids derived from pBIN-19, each containing a tandem repeat of one of the segments (C, M, N, R, S, U1, U2, U4). Each plasmid was transformed in one colony of *Agrobacterium tumefaciens* strain COR308.

The eight *A. tumefaciens* clones were grown separately in NZY+ medium (0.1% NZ amine, 0.5% yeast extract, 0.5% NaCl at pH7.5) completed with 12.5 mM MgCl₂, 12.5 mM MgSO₄ and 0.4% glucose, 10 mM MES pH 5.5, 50 µM acetosyringone, 50 mg.ml⁻¹ kanamycin, 25 mg.ml⁻¹ gentamycin and 5 mg.ml⁻¹ tetracycline, overnight at 28°C with stirring 150 rpm. Bacteria from each culture were then pelleted by centrifugation and resuspended in 5 mL of MS buffer (Murashige and Skoog, 1962) containing 1 mM MES (2-morpholin-4-ylethanesulfonic acid) pH 5.5 and 200 µM acetosyringone. The resuspensions were mixed

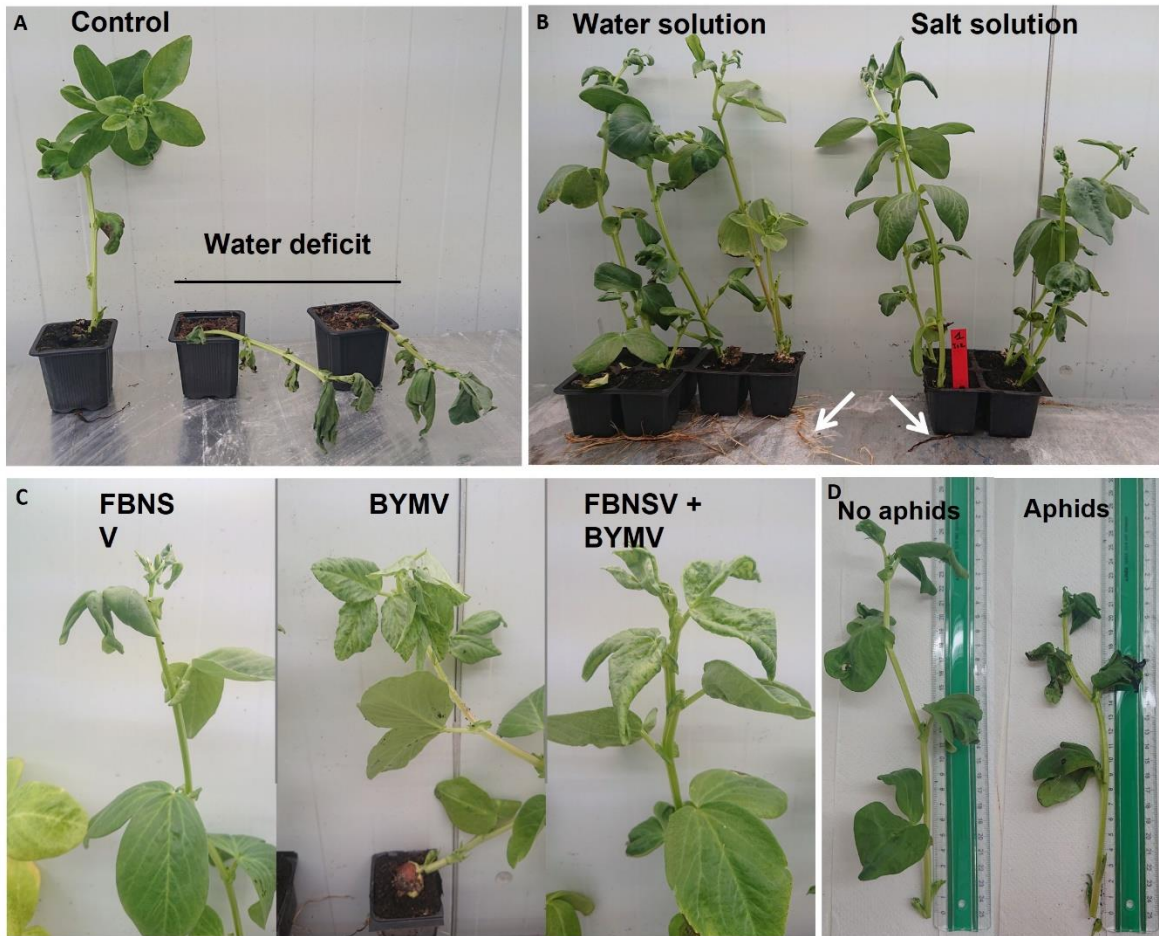


Figure 1: Symptoms of *V. faba* infected by the FBNSV and grown under different conditions: (A) water deficit, (B) salt excess, (C) BYMV co-infections or (D) aphid infestation. Plants were inoculated as described in the Methods section. Pictures were taken 3 weeks (A-C) or 13 days (D) after FBNSV inoculation. The white arrows in B highlight differences in root development that were consistently observed in the plant set with salt excess.

and incubated 2 h at room temperature. The bacterial mixture was inoculated by infiltration on the underside of young undeveloped leaves.

Experimental conditions

1- Water deficit

V. faba plants (var. Seville; Vilmorin) were grown in individual pots (6 cm x 6 cm x 6 cm) in a growth chamber with 13h/11h day/night photoperiod, 25/18°C day/night temperature. The seeds were watered only once abundantly after sowing to trigger germination. Nine days later, plantlets were agro-inoculated with FBNSV as described above. Plants grown in parallel under the same climatic conditions and watered regularly were used as controls. Total DNA was extracted 20 days post-inoculation (dpi) following the protocol described below. To facilitate tissue grinding and extraction, the plants were watered and thus rehydrated the day preceding extraction. A batch of non-rehydrated plants was also extracted to ensure that this additional watering did not immediately impact the genome formula. Physiological changes in plants caused by growing conditions, here water deprivation, were assessed by visual observations of severe wilting (**Figure 1A**).

2- Salt excess

V. faba plants were grown in a growth chamber with 13h/11h day/night photoperiod, 25/18°C day/night temperature. They were watered with a NaCl solution starting at 25 mM two days after sowing and increasing by 25 mM every two days up to a concentration of 100 mM maintained until the DNA extraction date. Plants watered with the same regime but with deionized water were used as controls. The seedlings were agro-inoculated nine days after sowing and total DNA extraction was done at 21 dpi. Physiological changes in plants caused by growing conditions, here salt-stress, were determined by visual observations of leaf color change (**Figure 1B**).

3- Co-infection with a potyvirus

The bean yellow mosaic virus (BYMV) infectious clone used in the co-inoculation experiments was provided by E. Johansen, University of Copenhagen, Denmark (Bruun-Rasmussen et al., 2007; Guyatt et al., 1996). Fifteen days-old *V. faba* were mechanically inoculated with a crude extract of BYMV infected tissues in sodium phosphate buffer (0.04 M) added with carborundum powder. FBNSV was inoculated when the light green mosaic symptoms of BYMV were visible (four or seven dpi). Two viruliferous aphids (*Acyrtosiphon pisum*, clone 210) that had acquired FBNSV during three days on infected plants were

deposited on each BYMV-infected recipient plant. Plants not infected with BYMV were inoculated with FBNSV in parallel to provide controls. After three days of inoculation access period, the aphids were eliminated by treatment with Pirimor. The plants were grown in a growth chamber with 13h/11h day/night photoperiod, 25/18°C day/night temperature. Total DNA was extracted from plants showing both FBNSV and BYMV symptoms (or FBNSV symptoms for controls) 21 days after FBNSV inoculation (**Figure 1C**).

4- Aphid infestation

Nine days old plantlets were agro-inoculated with FBNSV. The next day, ten aphids (*Aphis craccivora*) were placed on each of them. Plants without aphids grown in parallel were used as control. *V. faba* plants were grown in a growth chamber with 13h/11h day/night photoperiod, 25/18°C day/night temperature. Total DNA was extracted at 14 dpi when the aphid colony had become so large that it threatened the survival of the plants. The impact of aphid overcrowding on plant physiology and development was ascertained by visual observations (**Figure 1D**).

DNA extraction, qPCR quantification and data analysis

Total DNA was extracted from symptomatic apical leaves following the method described by (Di Mattia et al., 2020). The concentration of the total DNA extracted was estimated using a spectrophotometer NanoDrop™ 2000 (Thermo Scientific™, Waltham, MA, USA). The concentration of each FBNSV segment was estimated by qPCR using a LightCycler 480 thermocycler (Roche, Indianapolis, Ind, USA) as described by Bonnamy et al., 2023. Plants that did not test positive for all eight FBNSV segments were excluded from the analysis.

Total FBNSV concentration for each sample (virus titer) was obtained by adding the estimated concentrations of the eight segments and normalizing by the concentration of the total DNA extracted. Statistical analyses were done with ANOVA and Tukey HSD tests on log-transformed data to make their distribution closer to the Normal distribution and using the package “stats” of the software RStudio (<https://www.rstudio.com>).

The genome formula was calculated by expressing the concentration of each segment relatively to the total FBNSV concentration. Comparison of segment relative frequencies in the different conditions were performed with ANOVA and Tukey HSD tests on logit-transformed data to make their distribution closer to the Normal distribution and using the package “stats” of the software RStudio.

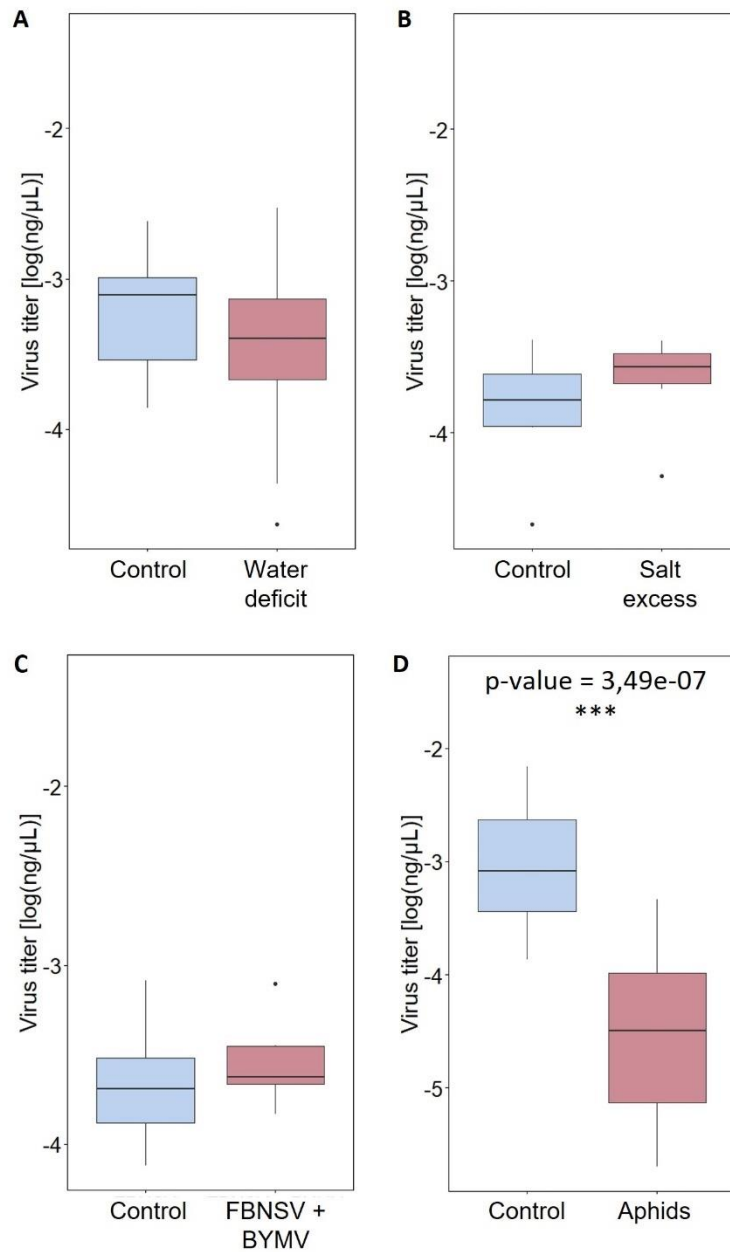


Figure 2: FBNSV viral loads in *V. faba* grown under different conditions: (A) water deficit, (B) salt excess, (C) BYMV co-infections or (D) aphid infestation (brown) compared to FBNSV viral load in standard growing conditions (blue). The central bar represents the median and the edges of the rectangle the first and third quartiles. The outer bars delineate 1.5 times the distance between the 25th and 75th percentile of the distribution. The dots represent the outliers. Statistically significant differences between conditions were assessed on log-transformed data based on ANOVA ($p \leq 0.05$) and Tukey HSD tests.

Results

1- Water deficit

Plants grown under water deficit conditions showed the first signs of wilting from around 5dpi and the wilting became increasingly severe until 19 dpi (**Figure 1A**). After one night of rehydration, the plants had recovered their turgor. Visual inspection indicated these plants were smaller than the regularly watered control plants, but no symptoms caused by FBNSV infection were visible on these plants before or after rehydration (**Figure 1**). The one-day rehydration had no significant impact on viral load or FBNSV genome formula in plant sets grown under water deprivation (**Table Sup. 1**). Despite the absence of virus-induced symptoms, the total concentration of FBNSV in plants grown in water deficit did not differ from that in regularly watered plants (**Figure 2**). In plants grown under water deficit, the relative frequency of segment M was statistically significantly lower (p-value = 0.0071) and that of U1 higher (p-value = 0.00157) than in control plants, but the frequency pattern was qualitatively conserved (**Figure 3; Table Sup. 2**).

2- Salt excess

Plants watered with saline solution had a similar size, but they consistently showed a reduced root system and a slightly darker leaf color when compared to control plants watered with deionized water (**Figure 1B**). Symptoms caused by FBNSV infection were similar to those of control plants (**Figure 1B**). The total accumulation of FBNSV also did not differ from controls (**Figure 2**). When analyzing and comparing the genome formula, the relative frequencies of segments S (p-value = 0.000113) and U4 (p-value = 0.0101) were statistically significantly lower in salt treated plants. Nevertheless, the frequency pattern of the eight segments remained qualitatively conserved (**Figure 3; Table Sup. 2**).

3- Co-infections with a potyvirus

Plants co-infected by both BYMV and FBNSV showed dwarfing and leaf curling typical of FBNSV infection and light green mosaic caused by BYMV infection (**Figure 1C**). Thus, co-infection results in an addition of symptoms of both diseases, with no noticeable synergism or antagonism. Consistently, the total concentration of FBNSV in the plants was similar whether BYMV was present or not (**Figure 2**). The genome formula in coinfecting plants did not statistically significantly differ from that in plants infected with FBNSV only, for none of the segments (**Figure 3; Table Sup. 2**), further confirming a complete absence of interference between the two viruses.

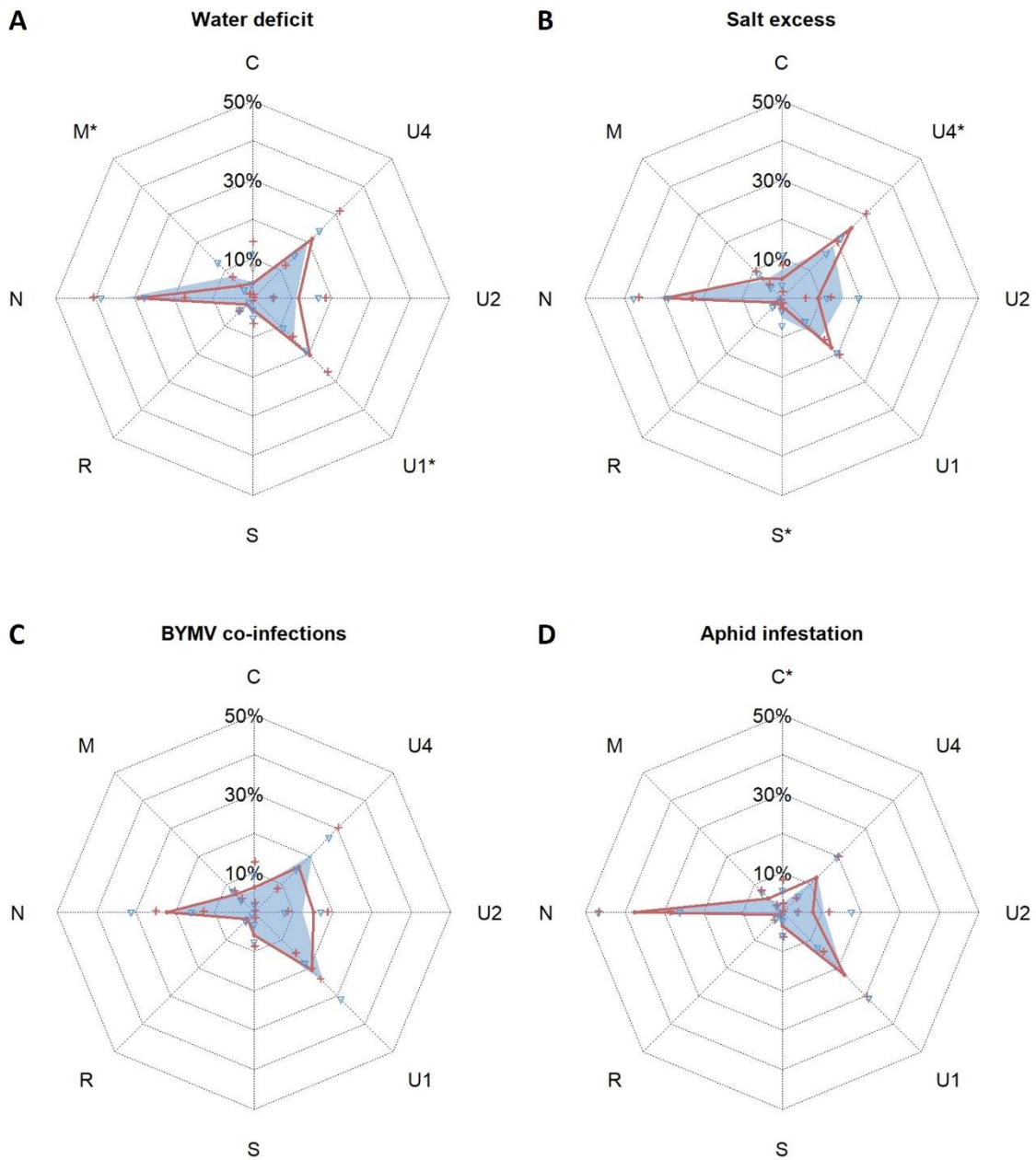


Figure 3: FBNSV genome formulas in *V. faba* grown under different conditions: (A) water deficit, (B) salt excess, (C) BYMV co-infections or (D) aphid infestation (brown) compared to plants grown under standard culture conditions (blue). Genome segment accumulation in symptomatic plants was estimated by qPCR and the relative frequency of the segments was determined. Standard deviations are represented by blue triangles or brown crosses. Asterisks associated to segment names indicate when the differences in frequencies between stressed and control plants are statistically significant (ANOVA ($p \leq 0.05$) and Tuckey HSD tests).

4- Aphid infestation

Aphid-infested plants were smaller than plants grown without aphids (**Figure 1**). All their leaves were curled inwards, whereas only the apex leaves were curled in on control plants (**Figure 1**). The total concentration of FBNSV in plants was significantly lower in aphid-infested plants compared to plants grown without aphids (**Figure 2**). Nevertheless, despite a statistically significant increase, relative to the control, in the frequency of segment C (p-value = 0.000753), the genome formulas of plants grown with and without aphids were very similar (**Figure 3; Table Sup. 2**).

Discussion and perspectives

To better appreciate the observations reported here, it is to be reminded that, when changing host species, the FBNSV genome formula can undergo massive changes. Indeed, in the example of FBNSV host switching from *V. faba* to *M. truncatula*, the relative frequency of almost all segments, except U1, was statistically significantly impacted. More importantly, the overall frequency pattern was drastically altered, with some segments that were very frequent in the first host becoming rare in the second, and *vice versa* (Sicard et al., 2013). For example, segment C represented 6% of the viral population in *V. faba* and increased to 33% when the infection was transferred to *M. truncatula*.

When compared to the situation earlier described (Sicard et al 2013, Gallet et al 2022) and summarized above, all culture conditions or stresses tested thus far do not have a major impact on the FBNSV genome formula. Some small, yet statistically significant variations are observed on some segments, depending on the stress, but the general pattern is not qualitatively affected.

Under conditions of aphid infestation, the total FBNSV titer is significantly reduced. This may be due to the plant's defenses against aphids, which also act on virus infection. Aphid feeding induces cytoskeletal rearrangement and callose deposition contributing to clogging the phloem sieve tubes (Czerniewicz et al., 2017; Zhao et al., 2016), which may hinder virus propagation. The decrease in viral load indicates that the FBNSV infection is altered by this plant condition, but it seems that no changes of the genome formula can overcome the problem.

For all other conditions tested, the total FBNSV load was no different from that in plants grown under standard conditions, suggesting that FBNSV infection is not significantly disturbed by the corresponding physiological status of the plant. That we in addition detected only very small modifications

of the genome formula can be interpreted in different ways: (i) the tested physiological conditions of the plant do not interfere in any way with the viral infection; (ii) small genome formula variations are sufficient to adjust the virus population accumulation to physiological changes in the host; (iii) the virus adjustments to the plant physiology do not proceed through genome formula modifications.

(i) The physiological condition of the plant may not interfere with FBNSV infection. This hypothesis cannot be excluded but appears poorly satisfactory. It has been shown that environmental constraints can have a direct impact on virus fitness and epidemiological parameters (González et al., 2020). Interferences due to crosstalk between abiotic and biotic stress (including viruses) response pathways have been evidenced (Fujita et al., 2006). More specifically, in the family *Geminiviridae*, the only other family of single-stranded DNA viruses infecting plants, the tomato yellow leaf curl virus (TYLCV) confers enhanced tolerance to drought (Mishra et al., 2022; Shteinberg et al., 2021) or high temperatures (Anfoka et al., 2016). The FBNSV being part of the related family *Nanoviridae*, a complete absence of crosstalk with abiotic stress pathways such as those mobilized under drought or heat would be surprising. In addition, interactions between geminivirus proteins and host proteins involved in stress response processes are identified. For example, the heat shock protein HSP70 which is involved in plant responses to heat and saline stresses, is also implicated in the translocation of the tomato yellow leaf curl Sardinia virus (TYLSV) coat protein (CP) from the cytoplasm to the nucleus of the host cell (Anfoka et al., 2016; Gorovits and Czosnek, 2017). The movement protein (MP) of the cabbage leaf curl virus (CaLCuV) interacts with the *Arabidopsis thaliana* synaptotagmin SYTA (Lewis and Lazarowitz, 2010), which has been shown to be involved in the cell-to-cell movement of various viruses via plasmodesmata (Uchiyama et al., 2014). SYTA belongs to the synaptotagmin family that are Ca²⁺/lipid binding proteins that includes Ca²⁺ sensors and is required for the maintenance of plasma membrane integrity and for tolerance to saline stress in *Arabidopsis thaliana* (Schapire et al., 2009; Yamazaki et al., 2009). Changes in the accumulation or availability of such proteins under stress conditions could force the virus to modify the expression of its genes to compensate for these modifications. In nanoviruses, it has been shown that the NSP protein encoded by segment N interacts with the host protein G3BP (Krapp et al., 2017). G3BP is a protein involved in the formation of stress granules, which are cellular structures that regulate protein expression during the stress response. NSP of FBNSV plays a role in aphid transmission but has no identified function in plants. Furthermore, the absence of segment N has no effect on FBNSV infection of *V. faba* under standard laboratory growth conditions. One could have expected a specific role of NSP in planta under stress conditions, but we did not observe noticeable changes in N frequency.

Besides the interactions between viral proteins and hosts, changes in cellular parameters generated by different culture conditions could have an impact on the production, stability or functionality of viral expression products. An increase in the copy number of the corresponding genes could compensate for this loss of functionality.

For all these reasons, it seems unlikely that the physiological state of the plant does not interfere with FBNSV infection. To find out whether the virus adjusts its gene expression to conditions generated by physiological changes in the plant, viral mRNAs should be quantified in the future.

(ii) Small variations may be sufficient to compensate for changes of the host physiological status. In the different conditions tested, we observed slight but statistically significant variations in the frequency of certain segments, which could somewhat be masked for some other segments by the inter-plant variability. Thus, we can speculate that small changes in gene copy number could be sufficient to adjust viral gene expression to a slightly modified environment. This contrasts with the drastic changes in the FBNSV genome formula between two host plants (Gallet et al., 2022; Sicard et al., 2013). Obviously, the genome formula is largely determined by the plant's genotype and much less so by its physiological state. It may be hypothesized that functional requirements differ or that certain viral proteins interact with different cellular factors depending on the host genotype, which would thus induce qualitative differences in the environment. This would require major changes in the expression of different viral genes from one host to another, and thus large variations in their copy number. Conversely, changes in physiological state could only represent quantitative changes, for example altering the amount of host proteins available, and small variations in the copy number of viral genes would suffice to rapidly adjust.

(iii) Finally, adjustments in FBNSV gene expression may not involve variations of the genome formula, but other mechanisms. Because of their high mutation rates, large population sizes and short generation times, viruses can adapt rapidly to adverse environments (González et al., 2020). Nanoviruses have particularly high mutation rates (Grigoras et al., 2010), potentially enabling rapid virus adaptation by selecting the variants best suited to new cellular conditions. However, the recent demonstration (Gallet et al 2022) that no mutations are responsible for the genome formula modification when the FBNSV switches from faba bean to alfalfa renders this hypothesis for immediate adaptation to stress conditions unlikely. An alternative possibility would be that viral promoters are differentially regulated under stress conditions, conferring a plasticity of the regulation of gene expression at the transcriptional level.

Finally, a major drawback of this work could well be that we have not exhaustively explored the possible physiological variations within the plant and that, eventually, one of the non yet tested conditions will impact indisputably the genome formula of FBNSV. To alleviate this drawback we are currently investigating a possible effect of the temperature (+/- 10°C) and rhizobial symbiosis to cumulate a biotic stress and nitrogen deprivation. These will enable us to cover a wider spectrum of conditions, and thereby finalize this body of work for publication .

References

- Abdelkhalek, A., El-Gendi, H., Al-Askar, A.A., Maresca, V., Moawad, H., Elsharkawy, M.M., Younes, H.A., Behiry, S.I., 2022. Enhancing systemic resistance in faba bean (*Vicia faba* L.) to Bean yellow mosaic virus via soil application and foliar spray of nitrogen-fixing *Rhizobium leguminosarum* bv. *viciae* strain 33504-Alex1. *Front. Plant Sci.* 13, 933498. <https://doi.org/10.3389/fpls.2022.933498>
- Andreola, S., Rodriguez, M., Parola, R., Alemano, S., Lascano, R., 2019. Interactions between soybean, *Bradyrhizobium japonicum* and Soybean mosaic virus: the effects depend on the interaction sequence. *Funct. Plant Biol.* 46, 1036. <https://doi.org/10.1071/FP17361>
- Anfoka, G., Moshe, A., Fridman, L., Amrani, L., Rotem, O., Kolot, M., Zeidan, M., Czosnek, H., Gorovits, R., 2016. Tomato yellow leaf curl virus infection mitigates the heat stress response of plants grown at high temperatures. *Sci. Rep.* 6, 19715. <https://doi.org/10.1038/srep19715>
- Bashir, S., Farrakh, S., Yasmin, T., Muhammad, A., Bashir, T., Manghwar, H., Mora-Poblete, F., Iqbal, S., Baazeem, A., Hyder, M.Z., 2022. Quantitation of Multipartite Banana Bunchy Top Virus Genomic Components and Their Transcripts in Infected Tissues of Banana (*Musa acuminata*). *Agronomy* 12, 2990. <https://doi.org/10.3390/agronomy12122990>
- Bayer, A., Brennan, G., Geballe, A.P., 2018. Adaptation by copy number variation in monopartite viruses. *Curr. Opin. Virol.* 33, 7–12. <https://doi.org/10.1016/j.coviro.2018.07.001>
- Berg, J., Willmann, S., Lässig, M., 2004. Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol.* 4, 42. <https://doi.org/10.1186/1471-2148-4-42>
- Bruun-Rasmussen, M., Møller, I.S., Tulinius, G., Hansen, J.K.R., Lund, O.S., Johansen, I.E., 2007. The Same Allele of Translation Initiation Factor 4E Mediates Resistance Against Two *Potyvirus* spp. in *Pisum sativum*. *Mol. Plant-Microbe Interactions*® 20, 1075–1082. <https://doi.org/10.1094/MPMI-20-9-1075>
- Czerniewicz, P., Sytykiewicz, H., Durak, R., Borowiak-Sobkowiak, B., Chrzanowski, G., 2017. Role of phenolic compounds during antioxidative responses of winter triticale to aphid and beetle attack. *Plant Physiol. Biochem.* 118, 529–540. <https://doi.org/10.1016/j.plaphy.2017.07.024>
- Di Mattia, J., Ryckebusch, F., Vernerey, M.-S., Piroilles, E., Sauvion, N., Peterschmitt, M., Zeddami, J.-L., Blanc, S., 2020. Co-Acquired Nanovirus and Geminivirus Exhibit a Contrasted Localization within Their Common Aphid Vector. *Viruses* 12, 299. <https://doi.org/10.3390/v12030299>
- Diefenbacher, M., Sun, J., Brooke, C.B., 2018. The parts are greater than the whole: the role of semi-infectious particles in influenza A virus biology. *Curr. Opin. Virol.* 33, 42–46. <https://doi.org/10.1016/j.coviro.2018.07.002>
- Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P., Malik, H.S., 2012. Poxviruses Deploy Genomic Accordions to Adapt Rapidly against Host Antiviral Defenses. *Cell* 150, 831–841. <https://doi.org/10.1016/j.cell.2012.05.049>
- Elliott, K.T., Cuff, L.E., Neidle, E.L., 2013. Copy number change: evolving views on gene amplification. *Future Microbiol.* 8, 887–899. <https://doi.org/10.2217/fmb.13.53>
- Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K., Shinozaki, K., 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr. Opin. Plant Biol.* 9, 436–442. <https://doi.org/10.1016/j.pbi.2006.05.014>
- Gallet, R., Di Mattia, J., Ravel, S., Zeddami, J.-L., Vitalis, R., Michalakakis, Y., Blanc, S., 2022. Gene copy number variations at the within-host population level modulate gene expression in a multipartite virus. *Virus Evol.* 8, veac058. <https://doi.org/10.1093/ve/veac058>

- González, R., Butković, A., Elena, S.F., 2020. From foes to friends: Viral infections expand the limits of host phenotypic plasticity, in: *Advances in Virus Research*. Elsevier, pp. 85–121. <https://doi.org/10.1016/bs.aivir.2020.01.003>
- Gorovits, R., Czosnek, H., 2017. The Involvement of Heat Shock Proteins in the Establishment of Tomato Yellow Leaf Curl Virus Infection. *Front. Plant Sci.* 8. <https://doi.org/10.3389/fpls.2017.00355>
- Grigoras, I., Timchenko, T., Grande-Pérez, A., Katul, L., Vetten, H.-J., Gronenborn, B., 2010. High Variability and Rapid Evolution of a Nanovirus. *J. Virol.* 84, 9105–9117. <https://doi.org/10.1128/JVI.00607-10>
- Grigoras, I., Timchenko, T., Katul, L., Grande-Pérez, A., Vetten, H.-J., Gronenborn, B., 2009. Reconstitution of Authentic Nanovirus from Multiple Cloned DNAs. *J. Virol.* 83, 10778–10787. <https://doi.org/10.1128/JVI.01212-09>
- Gronenborn, B., 2004. Nanoviruses: genome organisation and protein function. *Vet. Microbiol.* 98, 103–109. <https://doi.org/10.1016/j.vetmic.2003.10.015>
- Guillemaud, T., Raymond, M., Tsagkarakou, A., Bernard, C., Rochard, P., Pasteur, N., 1999. Quantitative variation and selection of esterase gene amplification in *Culex pipiens*. *Heredity* 83, 87–99. <https://doi.org/10.1038/sj.hdy.6885370>
- Gutiérrez, S., Zwart, M.P., 2018. Population bottlenecks in multicomponent viruses: first forays into the uncharted territory of genome-formula drift. *Curr. Opin. Virol.* 33, 184–190. <https://doi.org/10.1016/j.coviro.2018.09.001>
- Guyatt, K.J., Proll, D.F., Menssen, A., Davidson, A.D., 1996. The complete nucleotide sequence of bean yellow mosaic potyvirus RNA. *Arch. Virol.* 141, 1231–1246. <https://doi.org/10.1007/BF01718827>
- Guyot, V., Rajeswaran, R., Chu, H.C., Karthikeyan, C., Laboureau, N., Galzi, S., Mukwa, L.F.T., Krupovic, M., Kumar, P.L., Iskra-Caruana, M.-L., Pooggin, M.M., 2022. A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids. *PLOS Pathog.* 18, e1010448. <https://doi.org/10.1371/journal.ppat.1010448>
- Hu, Z., Zhang, X., Liu, W., Zhou, Q., Zhang, Q., Li, G., Yao, Q., 2016. Genome segments accumulate with different frequencies in *Bombyx mori bidensovirus*: The different frequencies of BmBDV genome segments. *J. Basic Microbiol.* 56, 1338–1343. <https://doi.org/10.1002/jobm.201600120>
- Jaouannet, M., Rodriguez, P.A., Thorpe, P., Lenoir, C.J.G., MacLeod, R., Escudero-Martinez, C., Bos, J.I.B., 2014. Plant immunity in plant aphid interactions. *Front. Plant Sci.* 5. <https://doi.org/10.3389/fpls.2014.00663>
- Krapp, S., Greiner, E., Amin, B., Sonnewald, U., Krenz, B., 2017. The stress granule component G3BP is a novel interaction partner for the nuclear shuttle proteins of the nanovirus pea necrotic yellow dwarf virus and geminivirus abutilon mosaic virus. *Virus Res.* 227, 6–14. <https://doi.org/10.1016/j.virusres.2016.09.021>
- Lal, A., Vo, T.T.B., Sanjaya, I.G.N.P.W., Ho, P.T., Kim, J.-K., Kil, E.-J., Lee, S., 2020. Nanovirus Disease Complexes: An Emerging Threat in the Modern Era. *Front. Plant Sci.* 11, 558403. <https://doi.org/10.3389/fpls.2020.558403>
- Lamers, J., Van Der Meer, T., Testerink, C., 2020. How Plants Sense and Respond to Stressful Environments. *Plant Physiol.* 182, 1624–1635. <https://doi.org/10.1104/pp.19.01464>
- Lewis, J.D., Knoblauch, M., Turgeon, R., 2022. The Phloem as an Arena for Plant Pathogens. *Annu. Rev. Phytopathol.* 60, 77–96. <https://doi.org/10.1146/annurev-phyto-020620-100946>
- Lewis, J.D., Lazarowitz, S.G., 2010. Arabidopsis synaptotagmin SYTA regulates endocytosis and virus movement protein cell-to-cell transport. *Proc. Natl. Acad. Sci.* 107, 2491–2496. <https://doi.org/10.1073/pnas.0909080107>
- Mansourpour, M., Gallet, R., Abbasi, A., Blanc, S., Dizadji, A., Zeddami, J.-L., 2022. Effects of an Alphasatellite on the Life Cycle of the Nanovirus *Faba Bean Necrotic Yellows Virus*. *J. Virol.* 96, e01388-21. <https://doi.org/10.1128/JVI.01388-21>

- Mascia, T., Gallitelli, D., 2016. Synergies and antagonisms in virus interactions. *Plant Sci.* 252, 176–192. <https://doi.org/10.1016/j.plantsci.2016.07.015>
- Medina-Ortega, K.J., Walker, G.P., 2015. Faba bean forisomes can function in defence against generalist aphids: Forisomes defend *V. faba* against generalist aphids. *Plant Cell Environ.* 38, 1167–1177. <https://doi.org/10.1111/pce.12470>
- Mishra, R., Shteinberg, M., Shkolnik, D., Anfoka, G., Czosnek, H., Gorovits, R., 2022. Interplay between abiotic (drought) and biotic (virus) stresses in tomato plants. *Mol. Plant Pathol.* 23, 475–488. <https://doi.org/10.1111/mpp.13172>
- Moreau, Y., Gil, P., Exbrayat, A., Rakotoarivony, I., Bréard, E., Sailleau, C., Viarouge, C., Zientara, S., Savini, G., Goffredo, M., Mancini, G., Loire, E., Gutierrez, S., 2020. The Genome Segments of Bluetongue Virus Differ in Copy Number in a Host-Specific Manner. *J. Virol.* 95, e01834-20. <https://doi.org/10.1128/JVI.01834-20>
- Murashige, T., Skoog, F., 1962. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant.* 15, 473–497. <https://doi.org/10.1111/j.1399-3054.1962.tb08052.x>
- Raymond, M., Chevillon, C., Guillemaud, T., Lenormand, T., Pasteur, N., 1998. An overview of the evolution of overproduced esterases in the mosquito *Culex pipiens*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 353, 1707–1711. <https://doi.org/10.1098/rstb.1998.0322>
- Revers, F., García, J.A., 2015. Molecular Biology of Potyviruses, in: *Advances in Virus Research*. Elsevier, pp. 101–199. <https://doi.org/10.1016/bs.aivir.2014.11.006>
- Schapiro, A.L., Voigt, B., Jasik, J., Rosado, A., Lopez-Cobollo, R., Menzel, D., Salinas, J., Mancuso, S., Valpuesta, V., Baluska, F., Botella, M.A., 2009. Arabidopsis Synaptotagmin 1 Is Required for the Maintenance of Plasma Membrane Integrity and Cell Viability. *Plant Cell* 20, 3374–3388. <https://doi.org/10.1105/tpc.108.063859>
- Shteinberg, M., Mishra, R., Anfoka, G., Altaieb, M., Brotman, Y., Moshelion, M., Gorovits, R., Czosnek, H., 2021. Tomato Yellow Leaf Curl Virus (TYLCV) Promotes Plant Tolerance to Drought. *Cells* 10, 2875. <https://doi.org/10.3390/cells10112875>
- Sicard, A., Michalakakis, Y., Gutiérrez, S., Blanc, S., 2016. The Strange Lifestyle of Multipartite Viruses. *PLOS Pathog.* 12, e1005819. <https://doi.org/10.1371/journal.ppat.1005819>
- Sicard, A., Yvon, M., Timchenko, T., Gronenborn, B., Michalakakis, Y., Gutierrez, S., Blanc, S., 2013. Gene copy number is differentially regulated in a multipartite virus. *Nat. Commun.* 4, 2248. <https://doi.org/10.1038/ncomms3248>
- Todd, R.T., Selmecki, A., 2020. Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. *eLife* 9, e58349. <https://doi.org/10.7554/eLife.58349>
- Tomanek, I., Grah, R., Lagator, M., Andersson, A.M.C., Bollback, J.P., Tkačik, G., Guet, C.C., 2020. Gene amplification as a form of population-level gene expression regulation. *Nat. Ecol. Evol.* 4, 612–625. <https://doi.org/10.1038/s41559-020-1132-7>
- Tuğrul, M., Paixão, T., Barton, N.H., Tkačik, G., 2015. Dynamics of Transcription Factor Binding Site Evolution. *PLOS Genet.* 11, e1005639. <https://doi.org/10.1371/journal.pgen.1005639>
- Uchiyama, A., Shimada-Beltran, H., Levy, A., Zheng, J.Y., Javia, P.A., Lazarowitz, S.G., 2014. The Arabidopsis synaptotagmin SYTA regulates the cell-to-cell movement of diverse plant viruses. *Front. Plant Sci.* 5. <https://doi.org/10.3389/fpls.2014.00584>
- Valli, A.A., Gallo, A., Rodamilans, B., López-Moya, J.J., García, J.A., 2018. The HCPro from the Potyviridae family: an enviable multitasking Helper Component that every virus would like to have: The HCPro from the Potyviridae family. *Mol. Plant Pathol.* 19, 744–763. <https://doi.org/10.1111/mpp.12553>
- Will, T., Tjallingii, W.F., Thönnessen, A., Van Bel, A.J.E., 2007. Molecular sabotage of plant defense by aphid saliva. *Proc. Natl. Acad. Sci.* 104, 10536–10541. <https://doi.org/10.1073/pnas.0703535104>

- Wu, B., Zwart, M.P., Sánchez-Navarro, J.A., Elena, S.F., 2017. Within-host Evolution of Segments Ratio for the Tripartite Genome of Alfalfa Mosaic Virus. *Sci. Rep.* 7, 5004. <https://doi.org/10.1038/s41598-017-05335-8>
- Yamazaki, T., Kawamura, Y., Minami, A., Uemura, M., 2009. Calcium-Dependent Freezing Tolerance in *Arabidopsis* Involves Membrane Resealing via Synaptotagmin SYT1. *Plant Cell* 20, 3389–3404. <https://doi.org/10.1105/tpc.108.062679>
- Yu, N.-T., Xie, H.-M., Zhang, Y.-L., Wang, J.-H., Xiong, Z., Liu, Z.-X., 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. <https://doi.org/10.1186/s12864-019-5901-0>
- Zhang, H., Zhu, J., Gong, Z., Zhu, J.-K., 2022. Abiotic stress responses in plants. *Nat. Rev. Genet.* 23, 104–119. <https://doi.org/10.1038/s41576-021-00413-0>
- Zhao, H., Sun, X., Xue, M., Zhang, X., Li, Q., 2016. Antioxidant Enzyme Responses Induced by Whiteflies in Tobacco Plants in Defense against Aphids: Catalase May Play a Dominant Role. *PLOS ONE* 11, e0165454. <https://doi.org/10.1371/journal.pone.0165454>
- Zhou, Y., Frey, T.K., Yang, J.J., 2009. Viral calciomics: Interplays between Ca²⁺ and virus. *Cell Calcium* 46, 1–17. <https://doi.org/10.1016/j.ceca.2009.05.005>
- Zwart, M.P., Elena, S.F., 2020. Modeling multipartite virus evolution: the genome formula facilitates rapid adaptation to heterogeneous environments[†]. *Virus Evol.* 6, veaa022. <https://doi.org/10.1093/ve/veaa022>

Supplemental Tables

Table Sup. 1: Statistical analysis of the impact of the one-day rehydration of infected plants on the FBNSV segment's relative frequency and total FBNSV titer

Comparisons were assessed through ANOVA and Tukey HSD tests on Log-transformed (FBNSV titer) or Logit-transformed (segment relative frequencies) data and using RStudio. dF indicates the number of observations and F-value the statistic value. No statistically significant differences ($p\text{-value} \leq 0.05$) were found.

FBNSV titer

df	25
F-value	1.091
p-value	0.306

Relative frequency of segments

Segment	C	M	N	R	S	U1	U2	U4
df	25							
F-value	0.352	0.701	1.141	0.631	0.494	2.605	0.052	0.761
p-value	0.707	0.506	0.336	0.54	0.616	0.0938	0.95	0.478

Table Sup. 2: Statistical analysis of the relative frequencies of each FBNSV segment in *V. faba* plants grown in different specific conditions compared to the genome formula in plants grown in standard conditions.

Comparisons of segment relative frequencies were assessed through ANOVA and Tukey HSD tests on Logit-transformed data and using RStudio. df indicates the number of observations and F-value the statistic value. The p-values indicating statistically significant differences (≤ 0.05) after FDR correction are in red.

Water deficit

Segment	C	M	N	R	S	U1	U2	U4
df	43							
F-value	0.491	7.994	2.473	0.654	2.184	11.4	0.026	0.061
p-value	0.487	0.0071	0.123	0.423	0.147	0.00157	0.874	0.806

Salt excess

Segment	C	M	N	R	S	U1	U2	U4
df	14							
F-value	0.131	0.215	0.028	1.73	28.04	1.424	2.639	8.841
p-value	0.723	0.65	0.87	0.21	0.000113	0.253	0.127	0.0101

BYMV co-infections

Segment	C	M	N	R	S	U1	U2	U4
df	38							
F-value	1.245	3.695	0.431	2.278	0.003	0.451	3.568	4.385
p-value	0.272	0.0621	0.515	0.139	0.956	0.506	0.0665	0.043

Aphid infestation

Segment	C	M	N	R	S	U1	U2	U4
df	58							
F-value	12.66	0.017	3.052	1.361	0.337	0.448	3.832	0.939
p-value	0.000753	0.897	0.0859	0.248	0.564	0.506	0.0551	0.337

Conclusion générale

Depuis leur découverte, les virus multipartites intriguent le monde scientifique par l'architecture et le packaging de leur génome. La mise au jour de la formule génomique par Sicard et al., 2013 a apporté encore un peu plus de mystère à ces virus. Cependant, l'intérêt porté à ce phénomène a augmenté ces dernières années, mettant en évidence des formules génomiques pour divers types de virus infectant différents règnes (Bashir et al., 2022; Diefenbacher et al., 2018; Guyot et al., 2022; Hu et al., 2016; Mansourpour et al., 2022; Moreau et al., 2020; Wu et al., 2017; Yu et al., 2019; Yvon et al., 2023). Ainsi, La formule génomique ne semble pas être une caractéristique des virus multipartites uniquement, mais plus largement des virus au génome segmenté.

Malgré cet engouement récent, les connaissances sur la formule génomique sont limitées et son importance pour le fonctionnement du virus reste à déterminer. Ainsi, il paraît primordial de répondre à deux grandes questions fondamentales sur la formule génomique. D'abord, la formule génomique a-t-elle un rôle fonctionnel ? Si certaines études le suggèrent, ceci doit encore être formellement démontré. Ensuite, par quel(s) mécanisme(s) est-elle mise en place et régulée ? C'est à cette question que ma thèse visait à apporter des réponses.

Nous avons déterminé que la formule génomique est en partie régulée au niveau du segment, certains segments (N et U4) ayant une plus grande capacité intrinsèque à s'accumuler au niveau local, probablement au niveau cellulaire. Cette plus grande accumulation peut s'expliquer soit par une plus faible dégradation de ces segments génomiques, soit par une aptitude supérieure à la réplication, soit par les deux. Les mécanismes de réplication des *Nanoviridae* ont été peu étudiés et sont majoritairement déduits des connaissances sur les *Geminiviridae*. Par ailleurs, même chez les géminivirus qui ont été extensivement étudiés, des points d'ombre subsistent et nécessitent que de nouvelles recherches soient menées afin de compléter les connaissances sur la réplication de ces virus. Ainsi, les données actuelles sur la réplication des *Nanoviridae* ne fournissent pas d'explication quant à la plus forte accumulation des segments N et U4 du FBNSV dans *V. faba*. Néanmoins, plusieurs pistes discutées dans la section « Concluding remarks » de l'article de revue devraient être explorées. Des efforts devraient être faits afin d'éclaircir le fonctionnement et le potentiel impact de la première étape de la réplication par RCR, l'initiation de la synthèse du brin complémentaire, dans l'accumulation des segments génomiques des *Nanoviridae*. Encore plus important, l'implication et le fonctionnement de la RDR dans la réplication de virus à ssDNA de plantes restent flous et aucune information concernant l'existence de ce mode de réplication chez les nanovirus n'existe.

Pourtant, ce mécanisme pourrait potentiellement contribuer à une réplication différentielle des segments génomiques.

Aussi, d'autres phénomènes, non testés dans cette thèse, pourraient impacter l'accumulation de chaque segment génomique de manière différente, par exemple une encapsidation ou une stabilité des particules virales différentes en fonction du segment génomique. L'hypothèse d'une encapsidation différentielle des segments peut difficilement être testée avec les outils dont nous disposons actuellement. Il aurait été intéressant de déterminer les ratios ADN encapsidé/ADN total pour chaque segment dans les systèmes d'infiltrations de feuilles ou de transfections de protoplastes, en infiltrant/transfectant chaque segment avec le segment R (pour la réplication) et le segment S (pour l'encapsidation). Dans le système d'infiltrations de feuilles, cela impliquerait que suffisamment de cellules soient co-infiltrées par les trois segments. Dans le cas du système de transfection, il faudrait pouvoir transfecter uniquement certains segments génomiques, ce qui n'est pas possible en utilisant une purification de particules virales. Enfin, la possibilité d'une stabilité de particules virales variables en fonction du segment qu'elles renferment est actuellement testée dans notre équipe afin de répondre à une autre problématique.

Outre la capacité intrinsèque des segments du FBNSV à s'accumuler localement, nous avons mis en évidence que la formule génomique est aussi façonnée par une régulation agissant au niveau du groupe de segments lors de l'infection systémique de la plante. L'absence de certains segments se répercute de manière inégale sur l'accumulation des autres segments. Nous n'avons pas constaté de cohérence entre la fonction du segment absent (lorsqu'elle est connue) et l'effet sur l'accumulation des autres segments, suggérant un fonctionnement complexe des composants viraux dans leur ensemble. Une fois encore, les segments N et U4 se démarquent des autres puisque leur absence n'affecte pas qualitativement la formule génomique. Ceci laisse supposer que leurs fréquences élevées dans la population virale est due à leur grande capacité à s'accumuler dans la plante par rapport aux autres segments. Dans leur ensemble, nos résultats indiquent que les deux niveaux de régulation ont tous les deux un rôle important dans la mise en place de la formule génomique. Le fonctionnement en groupe rend ce système encore plus intrigant. Le FBNSV est une entité fractionnée en plusieurs composants souvent localisés dans des cellules distinctes, mais qui évoluent ensemble pour façonner un système fonctionnel. A la différence d'un virus monopartite pour lequel la constitution du génome est bien délimitée, puisque celui-ci est contenu sur une seule molécule, les limites de notre système sont difficilement distinguables.

Notre étude n'apporte pas de réponse précise sur les mécanismes de régulation de la formule génomique, mais des indications sur les niveaux de cette régulation qui pourront servir de point de départ

à de nouvelles études. Par ailleurs, nos résultats confortent l'hypothèse d'une sélection du ou des groupe(s) de segments dont les fréquences relatives sont les plus optimales pour le fonctionnement du virus. Cette hypothèse suppose un rôle adaptatif de la formule génomique que nous ne sommes pas parvenus à démontrer. Les problèmes que nous avons rencontrés sont majoritairement d'ordre méthodologique, mais les différents concepts peuvent encore être exploités, soit en continuant la mise au point de protocoles, soit en utilisant de nouvelles approches.

Enfin, dans ce contexte d'une potentielle fonction adaptative, la formule génomique pourrait jouer un rôle dans régulation de l'expression génique par modulation du nombre de copies de gènes, en réponse à des changements environnementaux, comme un changement d'hôte (Gutiérrez and Zwart, 2018; Sicard et al., 2016, 2013; Zwart and Elena, 2020). Il a notamment été démontré, à chaque fois que cela a été recherché, que la formule génomique est déterminée par le génotype de l'hôte (Gallet et al., 2022; Moreau et al., 2020; Sicard et al., 2013; Wu et al., 2017). Cependant, aucune donnée n'existait concernant l'impact de l'état physiologique de la plante.

D'après nos premiers résultats, l'état physiologique de l'hôte n'entraîne pas de modification qualitative de la formule génomique, contrairement au génotype de celui-ci. Ceci est surprenant puisqu'il est fortement probable que les modifications de la physiologie interfèrent avec au moins une des étapes du cycle viral. Néanmoins, les légères variations de la fréquence observées pour certains segments pourraient révéler un léger ajustement du nombre de copie des gènes pour s'adapter au contexte cellulaire. A cause de la variabilité inter-plante de la formule génomique, nous ne pouvons, à ce stade, pas être certains que ces faibles variations de formule résultent réellement de l'effet des conditions de cultures. Ce point devra impérativement être clarifié. De plus, pour couvrir un spectre plus complet des différentes modifications physiologiques que peut subir la plante hôte, nous testons actuellement des conditions de culture complémentaires (différentes températures et symbiose rhizobienne).

Pour conclure, les études menées sur le FBNSV ces dernières années ont mis en lumière un fonctionnement original de ce virus à différents niveaux du cycle viral, renversant des dogmes de la virologie considérés comme universels. Le FBNSV constitue un bon modèle pour l'étude de la formule génomique. La nature ADN de son génome permet de dissocier les produits de réplication des produits d'expression. Le fait que chaque gène est porté par un segment différent facilite leur étude, permettant de les modifier ou de les retirer aisément. Néanmoins, la formule génomique reste tout de même un phénomène difficile à étudier.

Références hors articles

- Aronson, M.N., Meyer, A.D., Györgyey, J., Katul, L., Vetten, H.J., Gronenborn, B., Timchenko, T., 2000. Clink, a Nanovirus-Encoded Protein, Binds both pRB and SKP1. *J. Virol.* 74, 2967–2972. <https://doi.org/10.1128/JVI.74.7.2967-2972.2000>
- Baltimore, D., 1971. Expression of animal virus genomes. *Bacteriol Rev* 35, 235–241. <https://doi.org/10.1128/br.35.3.235-241.1971>
- Bashir, S., Farrakh, S., Yasmin, T., Muhammad, A., Bashir, T., Manghwar, H., Mora-Poblete, F., Iqbal, S., Baazeem, A., Hyder, M.Z., 2022. Quantitation of Multipartite Banana Bunchy Top Virus Genomic Components and Their Transcripts in Infected Tissues of Banana (*Musa acuminata*). *Agronomy* 12, 2990. <https://doi.org/10.3390/agronomy12122990>
- Bayer, A., Brennan, G., Geballe, A.P., 2018. Adaptation by copy number variation in monopartite viruses. *Current Opinion in Virology* 33, 7–12. <https://doi.org/10.1016/j.coviro.2018.07.001>
- Bechtold, N., Pelletier, G., 1998. In *Planta Agrobacterium Mediated Transformation of Adult Arabidopsis thaliana Plants by Vacuum Infiltration*, in: Martinez-Zapater, J.M., Salinas, J. (Eds.), *Arabidopsis Protocols*. Humana Press, Totowa, NJ, pp. 259–266. <https://doi.org/10.1385/0-89603-391-0:259>
- Betancourt, M., Fereres, A., Fraile, A., García-Arenal, F., 2008. Estimation of the Effective Number of Founders That Initiate an Infection after Aphid Transmission of a Multipartite Plant Virus. *J Virol* 82, 12416–12421. <https://doi.org/10.1128/JVI.01542-08>
- Blanc, S., Drucker, M., Uzest, M., 2014. Localizing Viruses in Their Insect Vectors. *Annu. Rev. Phytopathol.* 52, 403–425. <https://doi.org/10.1146/annurev-phyto-102313-045920>
- Briddon, R.W., Martin, D.P., Roumagnac, P., Navas-Castillo, J., Fiallo-Olivé, E., Moriones, E., Lett, J.-M., Zerbini, F.M., Varsani, A., 2018. Alphasatellitidae: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Arch Virol* 163, 2587–2600. <https://doi.org/10.1007/s00705-018-3854-2>
- Bruun-Rasmussen, M., Møller, I.S., Tulinius, G., Hansen, J.K.R., Lund, O.S., Johansen, I.E., 2007. The Same Allele of Translation Initiation Factor 4E Mediates Resistance Against Two *Potyvirus* spp. in *Pisum sativum*. *MPMI* 20, 1075–1082. <https://doi.org/10.1094/MPMI-20-9-1075>
- Burns, T.M., Harding, R.M., Dale, J.L., 1995. The genome organization of banana bunchy top virus: analysis of six ssDNA components. *Journal of General Virology* 76, 1471–1482. <https://doi.org/10.1099/0022-1317-76-6-1471>

- Cadena-Nava, R.D., Comas-Garcia, M., Garmann, R.F., Rao, A.L.N., Knobler, C.M., Gelbart, W.M., 2012. Self-Assembly of Viral Capsid Protein and RNA Molecules of Different Sizes: Requirement for a Specific High Protein/RNA Mass Ratio. *J Virol* 86, 3318–3326. <https://doi.org/10.1128/JVI.06566-11>
- Chao, L., 1991. Levels of selection, evolution of sex in RNA viruses, and the origin of life. *Journal of Theoretical Biology* 153, 229–246. [https://doi.org/10.1016/S0022-5193\(05\)80424-2](https://doi.org/10.1016/S0022-5193(05)80424-2)
- Choi, H., Jo, Y., Zhou, Y., Cho, W.K., 2019. First Report of Milk Vetch Dwarf Virus Infecting Lily in Korea. *Plant Disease* 103, 2144. <https://doi.org/10.1094/PDIS-02-19-0282-PDN>
- Chu, P.W., Boevink, P., Surin, B., Larkin, P., Keese, P., Waterhouse, P.M., 1995. Non-germinated single-stranded DNA plant viruses. *Pathogenesis and Host Specificity in Plant Disease, Volume III: Viruses & Viroids* 311-341.
- Clough, S.J., Bent, A.F., 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*: Floral dip transformation of *Arabidopsis*. *The Plant Journal* 16, 735–743. <https://doi.org/10.1046/j.1365-313x.1998.00343.x>
- Dall’Ara, M., Ratti, C., Bouzoubaa, S., Gilmer, D., 2016. Ins and Outs of Multipartite Positive-Strand RNA Plant Viruses: Packaging versus Systemic Spread. *Viruses* 8, 228. <https://doi.org/10.3390/v8080228>
- Di Mattia, J., Ryckebusch, F., Vernerey, M.-S., Pirolles, E., Sauvion, N., Peterschmitt, M., Zeddami, J.-L., Blanc, S., 2020a. Co-Acquired Nanovirus and Geminivirus Exhibit a Contrasted Localization within Their Common Aphid Vector. *Viruses* 12, 299. <https://doi.org/10.3390/v12030299>
- Di Mattia, J., Torralba, B., Yvon, M., Zeddami, J.-L., Blanc, S., Michalakakis, Y., 2022. Nonconcomitant host-to-host transmission of multipartite virus genome segments may lead to complete genome reconstitution. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2201453119. <https://doi.org/10.1073/pnas.2201453119>
- Di Mattia, J., Vernerey, M.-S., Yvon, M., Pirolles, E., Villegas, M., Gaafar, Y., Ziebell, H., Michalakakis, Y., Zeddami, J.-L., Blanc, S., 2020b. Route of a Multipartite Nanovirus across the Body of Its Aphid Vector. *J Virol* 94, e01998-19. <https://doi.org/10.1128/JVI.01998-19>
- Di Mattia, J., Zeddami, J.-L., Uzest, M., Blanc, S., 2023. The helper strategy in vector-transmission of plant viruses. *Peer Community Journal* 3, e32. <https://doi.org/10.24072/pcjournal.258>
- Diefenbacher, M., Sun, J., Brooke, C.B., 2018. The parts are greater than the whole: the role of semi-infectious particles in influenza A virus biology. *Current Opinion in Virology* 33, 42–46. <https://doi.org/10.1016/j.coviro.2018.07.002>

- Dietzgen, R., Mann, K., Johnson, K., 2016. Plant Virus–Insect Vector Interactions: Current and Potential Future Research Directions. *Viruses* 8, 303. <https://doi.org/10.3390/v8110303>
- Dugdale, B., Becker, D.K., Beetham, P.R., Harding, R.M., Dale, J.L., 2000. Promoters derived from banana bunchy top virus DNA-1 to –5 direct vascular-associated expression in transgenic banana (*Musa* spp.). *Plant Cell Reports* 19, 810–814. <https://doi.org/10.1007/s002999900185>
- Dugdale, B., Beetham, P.R., Becker, D.K., Harding, R.M., Dale, J.L., 1998. Promoter activity associated with the intergenic regions of banana bunchy top virus DNA-1 to -6 in transgenic tobacco and banana cells. *Journal of General Virology*, 79, 2301–2311. <https://doi.org/10.1099/0022-1317-79-10-2301>
- Elde, N.C., Child, S.J., Eickbush, M.T., Kitzman, J.O., Rogers, K.S., Shendure, J., Geballe, A.P., Malik, H.S., 2012. Poxviruses Deploy Genomic Accordions to Adapt Rapidly against Host Antiviral Defenses. *Cell* 150, 831–841. <https://doi.org/10.1016/j.cell.2012.05.049>
- Fauquet, C.M., 1999. TAXONOMY, CLASSIFICATION AND NOMENCLATURE OF VIRUSES, in: *Encyclopedia of Virology*. Elsevier, pp. 1730–1756. <https://doi.org/10.1006/rwvi.1999.0277>
- Feng, J.-L., Chen, S.-N., Tang, X.-S., Ding, X.-F., Du, Z.-Y., Chen, J.-S., 2006. Quantitative Determination of *Cucurbit Mosaic Virus* Genome RNAs in Virions by Real-Time Reverse Transcription-Polymerase Chain Reaction. *ABBS* 38, 669–676. <https://doi.org/10.1111/j.1745-7270.2006.00216.x>
- Franz, A.W.E., Van Der Wilk, F., Verbeek, M., Dullemans, A.M., Van Den Heuvel, J.F.J.M., 1999. Faba Bean Necrotic Yellows Virus (Genus *Nanovirus*) Requires a Helper Factor for Its Aphid Transmission. *Virology* 262, 210–219. <https://doi.org/10.1006/viro.1999.9904>
- French, R., Ahlquist, P., 1988. Characterization and engineering of sequences controlling in vivo synthesis of brome mosaic virus subgenomic RNA. *J Virol* 62, 2411–2420. <https://doi.org/10.1128/jvi.62.7.2411-2420.1988>
- Fulton, R.W., 1980. Biological Significance of Multicomponent Viruses. *Annu. Rev. Phytopathol.* 18, 131–146. <https://doi.org/10.1146/annurev.py.18.090180.001023>
- Gallet, R., Di Mattia, J., Ravel, S., Zeddami, J.-L., Vitalis, R., Michalakakis, Y., Blanc, S., 2022. Gene copy number variations at the within-host population level modulate gene expression in a multipartite virus. *Virus Evolution* 8, veac058. <https://doi.org/10.1093/ve/veac058>
- Gallet, R., Fabre, F., Thébaud, G., Sofonea, M.T., Sicard, A., Blanc, S., Michalakakis, Y., 2018. Small Bottleneck Size in a Highly Multipartite Virus during a Complete Infection Cycle. *J Virol* 92, e00139-18. <https://doi.org/10.1128/JVI.00139-18>

- García-Arriaza, J., Manrubia, S.C., Toja, M., Domingo, E., Escarmís, C., 2004. Evolutionary Transition toward Defective RNAs That Are Infectious by Complementation. *J Virol* 78, 11678–11685. <https://doi.org/10.1128/JVI.78.21.11678-11685.2004>
- Gilmer, D., Ratti, C., Michel, F., 2018. Long-distance movement of helical multipartite phytoviruses: keep connected or die? *Current Opinion in Virology* 33, 120–128. <https://doi.org/10.1016/j.coviro.2018.07.016>
- González-Jara, P., Fraile, A., Canto, T., García-Arenal, F., 2009. The Multiplicity of Infection of a Plant Virus Varies during Colonization of Its Eukaryotic Host. *J Virol* 83, 7487–7494. <https://doi.org/10.1128/JVI.00636-09>
- Grigoras, I., Ginzo, A.I.D.C., Martin, D.P., Varsani, A., Romero, J., Mammadov, A.Ch., Huseynova, I.M., Aliyev, J.A., Kheyir-Pour, A., Huss, H., Ziebell, H., Timchenko, T., Vetten, H.-J., Gronenborn, B., 2014. Genome diversity and evidence of recombination and reassortment in nanoviruses from Europe. *Journal of General Virology* 95, 1178–1191. <https://doi.org/10.1099/vir.0.063115-0>
- Grigoras, I., Timchenko, T., Katul, L., Grande-Pérez, A., Vetten, H.-J., Gronenborn, B., 2009. Reconstitution of Authentic Nanovirus from Multiple Cloned DNAs. *J Virol* 83, 10778–10787. <https://doi.org/10.1128/JVI.01212-09>
- Grigoras, I., Vetten, H.-J., Commandeur, U., Ziebell, H., Gronenborn, B., Timchenko, T., 2018. Nanovirus DNA-N encodes a protein mandatory for aphid transmission. *Virology* 522, 281–291. <https://doi.org/10.1016/j.virol.2018.07.001>
- Gutiérrez, S., Michalakis, Y., Blanc, S., 2012. Virus population bottlenecks during within-host progression and host-to-host transmission. *Current Opinion in Virology* 2, 546–555. <https://doi.org/10.1016/j.coviro.2012.08.001>
- Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y., Blanc, S., 2010. Dynamics of the Multiplicity of Cellular Infection in a Plant Virus. *PLoS Pathog* 6, e1001113. <https://doi.org/10.1371/journal.ppat.1001113>
- Gutiérrez, S., Zwart, M.P., 2018. Population bottlenecks in multicomponent viruses: first forays into the uncharted territory of genome-formula drift. *Current Opinion in Virology* 33, 184–190. <https://doi.org/10.1016/j.coviro.2018.09.001>
- Guyatt, K.J., Proll, D.F., Menssen, A., Davidson, A.D., 1996. The complete nucleotide sequence of bean yellow mosaic potyvirus RNA. *Archives of Virology* 141, 1231–1246. <https://doi.org/10.1007/BF01718827>

- Guyot, V., Rajeswaran, R., Chu, H.C., Karthikeyan, C., Laboureau, N., Galzi, S., Mukwa, L.F.T., Krupovic, M., Kumar, P.L., Iskra-Caruana, M.-L., Pooggin, M.M., 2022. A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids. *PLoS Pathog* 18, e1010448. <https://doi.org/10.1371/journal.ppat.1010448>
- Hajimorad, M.R., Kurath, G., Randles, J.W., Francki, R.I.B., 1991. Change in Phenotype and Encapsidated RNA Segments of an Isolate of Alfalfa Mosaic Virus: An Influence of Host Passage. *Journal of General Virology* 72, 2885–2893. <https://doi.org/10.1099/0022-1317-72-12-2885>
- Harrison, B.D., Kubo, S., Robinson, D.J., Hutcheson, A.M., 1976. The Multiplication Cycle of Tobacco Rattle Virus in Tobacco Mesophyll Protoplasts. *Journal of General Virology* 33, 237–248. <https://doi.org/10.1099/0022-1317-33-2-237>
- Hasanvand, V., Heydanejad, J., Massumi, H., Kleinow, T., Jeske, H., Fontenele, R.S., 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>
- Horser, C.L., Harding, R.M., Dale, J.L., 2001. Banana bunchy top nanovirus DNA-1 encodes the ‘master’ replication initiation protein. *Journal of General Virology* 82, 459–464. <https://doi.org/10.1099/0022-1317-82-2-459>
- Hu, Z., Zhang, X., Liu, W., Zhou, Q., Zhang, Q., Li, G., Yao, Q., 2016. Genome segments accumulate with different frequencies in *Bombyx mori bidensovirus*: The different frequencies of BmBDV genome segments. *J. Basic Microbiol.* 56, 1338–1343. <https://doi.org/10.1002/jobm.201600120>
- Iranzo, J., Manrubia, S.C., 2012. Evolutionary dynamics of genome segmentation in multipartite viruses. *Proc. R. Soc. B.* 279, 3812–3819. <https://doi.org/10.1098/rspb.2012.1086>
- Ji, X., Yu, N., Qu, L., Li, B., Liu, Z., 2019. Banana bunchy top virus (BBTV) nuclear shuttle protein interacts and re-distributes BBTV coat protein in *Nicotiana benthamiana*. *3 Biotech* 9, 121. <https://doi.org/10.1007/s13205-019-1656-1>
- Johansen, I.E., Keller, K.E., Dougherty, W.G., Hampton, R.O., 1996. Biological and molecular properties of a pathotype P-1 and a pathotype P-4 isolate of pea seed-borne mosaic virus. *Journal of General Virology* 77, 1329–1333. <https://doi.org/10.1099/0022-1317-77-6-1329>
- Johansen, I.E., Lund, O.S., Hjulsager, C.K., Laursen, J., 2001. Recessive resistance in *Pisum sativum* and potyvirus pathotype resolved in a gene-for-cistron correspondence between host and virus. *J. Virol.* 75, 6609–6614. <https://doi.org/10.1128/JVI.75.14.6609-6614.2001>

- Katul, L., Maiss, E., Morozov, S.Y., Vetten, H.J., 1997. Analysis of six DNA components of the faba bean necrotic yellows virus genome and their structural affinity to related plant virus genomes. *Virology* 233, 247–259. <https://doi.org/10.1006/viro.1997.8611>
- Krenz, B., Schießl, I., Greiner, E., Krapp, S., 2017. Analyses of pea necrotic yellow dwarf virus-encoded proteins. *Virus Genes* 53, 454–463. <https://doi.org/10.1007/s11262-017-1439-x>
- Lal, A., Kil, E., Thuy, V.T.B., Fadhila, C., Ho, P.T., Byun, H., Dao, H.T., Kim, J., Lee, S., 2020a. Milk vetch dwarf virus infection in the Solanaceae and Caricaceae families in Southeast Asia. *Plant Pathol* 69, 1026–1033. <https://doi.org/10.1111/ppa.13196>
- Lal, A., Kil, E.-J., Byun, H.-S., Zarghami D., S., Kim, J.-K., Lee, S., 2018. First Report of *Milk vetch dwarf virus* Associated with Dwarfism in Papaya in Korea. *Plant Disease* 102, 2666–2666. <https://doi.org/10.1094/PDIS-01-18-0190-PDN>
- Lal, A., Vo, T.T.B., Sanjaya, I.G.N.P.W., Ho, P.T., Kim, J.-K., Kil, E.-J., Lee, S., 2020b. Nanovirus Disease Complexes: An Emerging Threat in the Modern Era. *Front. Plant Sci.* 11, 558403. <https://doi.org/10.3389/fpls.2020.558403>
- Leeks, A., Young, P.G., Turner, P.E., Wild, G., West, S.A., 2023. Cheating leads to the evolution of multipartite viruses. *PLoS Biol* 21, e3002092. <https://doi.org/10.1371/journal.pbio.3002092>
- Lucía-Sanz, A., Aguirre, J., Manrubia, S., 2018. Theoretical approaches to disclosing the emergence and adaptive advantages of multipartite viruses. *Current Opinion in Virology* 33, 89–95. <https://doi.org/10.1016/j.coviro.2018.07.018>
- Lucía-Sanz, A., Manrubia, S., 2017. Multipartite viruses: adaptive trick or evolutionary treat? *npj Syst Biol Appl* 3, 34. <https://doi.org/10.1038/s41540-017-0035-y>
- 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*. *J Virol* 96, e01388-21. <https://doi.org/10.1128/JVI.01388-21>
- Michalakis, Y., Blanc, S., 2020. The Curious Strategy of Multipartite Viruses. *Annu. Rev. Virol.* 7, 203–218. <https://doi.org/10.1146/annurev-virology-010220-063346>
- Miyashita, S., Kishino, H., 2010. Estimation of the Size of Genetic Bottlenecks in Cell-to-Cell Movement of *Soil-Borne Wheat Mosaic Virus* and the Possible Role of the Bottlenecks in Speeding Up Selection of Variations in *trans*-Acting Genes or Elements. *J Virol* 84, 1828–1837. <https://doi.org/10.1128/JVI.01890-09>

- Morante-Carriel, J., Sellés-Marchart, S., Martínez-Márquez, A., Martínez-Esteso, M.J., Luque, I., Bru-Martínez, R., 2014. RNA isolation from loquat and other recalcitrant woody plants with high quality and yield. *Analytical Biochemistry* 452, 46–53. <https://doi.org/10.1016/j.ab.2014.02.010>
- Moreau, Y., Gil, P., Exbrayat, A., Rakotoarivony, I., Bréard, E., Sailleau, C., Viarouge, C., Zientara, S., Savini, G., Goffredo, M., Mancini, G., Loire, E., Gutierrez, S., 2020. The Genome Segments of Bluetongue Virus Differ in Copy Number in a Host-Specific Manner. *J Virol* 95, e01834-20. <https://doi.org/10.1128/JVI.01834-20>
- Naderpour, M., Johansen, I.E., 2011. Visualization of resistance responses in *Phaseolus vulgaris* using reporter tagged clones of Bean common mosaic virus. *Virus Research* 159, 1–8. <https://doi.org/10.1016/j.virusres.2011.04.004>
- Nee, S., 1987. The evolution of multicompartmental genomes in viruses. *J Mol Evol* 25, 277–281. <https://doi.org/10.1007/BF02603110>
- Ojosnegros, S., García-Arriaza, J., Escarmís, C., Manrubia, S.C., Perales, C., Arias, A., Mateu, M.G., Domingo, E., 2011. Viral Genome Segmentation Can Result from a Trade-Off between Genetic Content and Particle Stability. *PLoS Genet* 7, e1001344. <https://doi.org/10.1371/journal.pgen.1001344>
- Park, H., Denha, S., Higgs, P.G., 2023. Evolution of Bipartite and Segmented Viruses from Monopartite Viruses. *Viruses* 15, 1135. <https://doi.org/10.3390/v15051135>
- Pressing, J., Reaney, D.C., 1984. Divided genomes and intrinsic noise. *J Mol Evol* 20, 135–146. <https://doi.org/10.1007/BF02257374>
- Qazi, J., 2016. Banana bunchy top virus and the bunchy top disease. *J Gen Plant Pathol* 82, 2–11. <https://doi.org/10.1007/s10327-015-0642-7>
- Sanderfoot, A.A., Lazarowitz, S.G., 1995. Cooperation in Viral Movement: The Geminivirus BL1 Movement Protein Interacts with BR1 and Redirects It from the Nucleus to the Cell Periphery. *Plant Cell* 1185–1194. <https://doi.org/10.1105/tpc.7.8.1185>
- Sanjuán, R., Thoulouze, M.-I., 2019. Why viruses sometimes disperse in groups? *Virus Evolution* 5, vez025. <https://doi.org/10.1093/ve/vez025>
- Shirasawa-Seo, N., Sano, Y., Nakamura, S., Murakami, T., Seo, S., Ohashi, Y., Hashimoto, Y., Matsumoto, T., 2005. Characteristics of the promoters derived from the single-stranded DNA components of Milk vetch dwarf virus in transgenic tobacco. *Journal of General Virology*, 86, 1851–1860. <https://doi.org/10.1099/vir.0.80790-0>

- Sicard, A., 2014. Fonctionnement des populations de virus multipartites de plantes au cours des différentes étapes de leur cycle de vie.
- Sicard, A., Michalakakis, Y., Gutiérrez, S., Blanc, S., 2016. The Strange Lifestyle of Multipartite Viruses. *PLoS Pathog* 12, e1005819. <https://doi.org/10.1371/journal.ppat.1005819>
- Sicard, A., Piroles, E., Gallet, R., Vernerey, M.-S., Yvon, M., Urbino, C., Peterschmitt, M., Gutierrez, S., Michalakakis, Y., Blanc, S., 2019. A multicellular way of life for a multipartite virus. *eLife* 8, e43599. <https://doi.org/10.7554/eLife.43599>
- Sicard, A., Yvon, M., Timchenko, T., Gronenborn, B., Michalakakis, Y., Gutierrez, S., Blanc, S., 2013. Gene copy number is differentially regulated in a multipartite virus. *Nat Commun* 4, 2248. <https://doi.org/10.1038/ncomms3248>
- Sicard, A., Zeddiam, J.-L., Yvon, M., Michalakakis, Y., Gutiérrez, S., Blanc, S., 2015. Circulative Nonpropagative Aphid Transmission of Nanoviruses: an Oversimplified View. *J Virol* 89, 9719–9726. <https://doi.org/10.1128/JVI.00780-15>
- Skalka, A.M., Flint, J., Rall, G.F., Racaniello, V.R., 2015. Principles of Virology, Volume I: Molecular Biology. American Society of Microbiology. <https://doi.org/10.1128/9781555818951>
- Thomas, J.E., Gronenborn, B., Harding, R.M., Mandal, B., Grigoras, I., Randles, J.W., Sano, Y., Timchenko, T., Vetten, H.J., Yeh, H.-H., Ziebell, H., ICTV Report Consortium, 2021. ICTV Virus Taxonomy Profile: Nanoviridae. *Journal of General Virology* 102. <https://doi.org/10.1099/jgv.0.001544>
- Timchenko, T., Katul, L., Aronson, M., Vega-Arreguín, J.C., Ramirez, B.C., Vetten, H.J., 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. <https://doi.org/10.1099/vir.0.81753-0>
- Timchenko, T., Katul, L., Sano, Y., de Kouchkovsky, F., Vetten, H.J., 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. <https://doi.org/10.1006/viro.2000.0439>
- Timchenko, T., Kouchkovsky, F. de, Katul, L., David, C., Vetten, H.J., 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.
- Tomanek, I., Grah, R., Lagator, M., Andersson, A.M.C., Bollback, J.P., Tkačik, G., Guet, C.C., 2020. Gene amplification as a form of population-level gene expression regulation. *Nat Ecol Evol* 4, 612–625. <https://doi.org/10.1038/s41559-020-1132-7>

- Trapani, S., Bhat, E.A., Yvon, M., Lai-Kee-Him, J., Hoh, F., Vernerey, M.-S., Pirolles, E., Bonnamy, M., Schoehn, G., Zeddami, J.-L., Blanc, S., Bron, P., 2023. Structure-guided mutagenesis of the capsid protein indicates that a nanovirus requires assembled viral particles for systemic infection. *PLoS Pathog* 19, e1011086. <https://doi.org/10.1371/journal.ppat.1011086>
- Tomas, N., Zwart, M.P., Lafforgue, G., Elena, S.F., 2014. Within-Host Spatiotemporal Dynamics of Plant Virus Infection at the Cellular Level. *PLoS Genet* 10, e1004186. <https://doi.org/10.1371/journal.pgen.1004186>
- Vega-Arreguín, J.C., Gronenborn, B., Ramírez, B.C., 2007. *Arabidopsis thaliana* is a host of the legume nanovirus Faba bean necrotic yellows virus. *Virus Research* 128, 81–87. <https://doi.org/10.1016/j.virusres.2007.04.011>
- Vetten, H.J., Knierim, D., Rakoski, M.S., Menzel, W., Maiss, E., Gronenborn, B., Winter, S., Krenz, B., 2019. Identification of a novel nanovirus in parsley. *Arch Virol* 164, 1883–1887. <https://doi.org/10.1007/s00705-019-04280-3>
- Wanitchakorn, R., Hafner, G.J., Harding, R.M., Dale, J.L., 2000. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *Microbiology* 81, 299–306. <https://doi.org/10.1099/0022-1317-81-1-299>
- Wanitchakorn, R., Harding, R.M., Dale, J.L., 1997. Banana bunchy top virus DNA-3 encodes the viral coat protein. *Arch. Virol.* 142, 1673–1680. <https://doi.org/10.1007/s007050050188>
- Wu, B., Zwart, M.P., Sánchez-Navarro, J.A., Elena, S.F., 2017. Within-host Evolution of Segments Ratio for the Tripartite Genome of Alfalfa Mosaic Virus. *Sci Rep* 7, 5004. <https://doi.org/10.1038/s41598-017-05335-8>
- Yu, N.-T., Xie, H.-M., Zhang, Y.-L., Wang, J.-H., Xiong, Z., Liu, Z.-X., 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. <https://doi.org/10.1186/s12864-019-5901-0>
- Zwart, M.P., Elena, S.F., 2020. Modeling multipartite virus evolution: the genome formula facilitates rapid adaptation to heterogeneous environments†. *Virus Evolution* 6, veaa022. <https://doi.org/10.1093/ve/veaa022>
- Zwart, M.P., Elena, S.F., 2015. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annu. Rev. Virol.* 2, 161–179. <https://doi.org/10.1146/annurev-virology-100114-055135>

Annexe 1

Protocole de protoplastation, transfection et culture

Préparation des solutions :

Solution enzymatique

pour 100 mL

Mannitol 0,5 M 7,28 g

Cellulase R10 0,5 % 1,5 g

Macérozyme R10 0,4 g

KCl 20 mM 0,149 g

MES 1 M 0,39 g

Ajouter 90 mL d'eau ; ajuster le pH à 5,7 avec du KOH 1 N

Filtrer avec un filtre 45 µm

CaCl₂ 0,146 g

BSA 1 % 0,1 g

Conserver à -20 °C

Solution W5

pour 200 mL

NaCl 1,8 g

CaCl₂ 3,67 g

KCl 0,074 g

MES 0,078 g

180 mL ; ajuster le pH à 5,7 ; compléter à 200 mL

Stériliser par filtration et conserver à 4°C

Solution MMg

pour 100 mL

Mannitol 7,8 g

MgCl₂ 0,3 g

MES 0,039 g

80 mL d'eau ; ajuster le pH à 5,7 ; compléter avec de l'eau jusqu'à 100 mL

Stériliser par filtration et conserver à 4°C

Solution PEG

pour 10 mL

PEG3600 4 g

Mannitol 0,364 g

CaCl₂ 0,146 g

Ajuster avec de l'eau à 10 mL

Stériliser par autoclavage et stocker à -20°C

Milieu Gamborg complété

pour 50 mL

MES (1M) 250 µL

Mannitol 4,5 g

Stériliser par filtration et conserver à 4°C

Protoplastation :

- Stériliser les feuilles (plante de 12 jours) au Domestos (1 mL + H₂O QSP 25 mL) pendant 1 min puis 4 rinçages successifs à l'eau milliRo stérile
- Dans une boîte de pétri, infiltrer puis incubé les feuilles avec 14 mL de la solution enzymatique
- Parafilmer et envelopper dans du papier aluminium pour protéger de la lumière
- Incuber environ 3h15 à 28°C avec agitation rotative 50 rpm/min
- Transférer la solution de protoplastes dans des tubes de 10 mL à fond rond
- Centrifuger 4 min à 80 g ; accélération 5, décélération 1

Sur glace :

- Retirer le surnageant et re-suspendre le culot dans 8 mL de solution W5 froide
- Centrifuger 5 min à 4 °C, 80 g ; accélération 5, décélération 1
- Re-suspendre le culot dans 5 mL de solution W5 froide ; sortir la sol W5 de la glace
- Laisser les protoplastes sur glace pendant 30 min
- Déterminer la concentration des protoplastes : 30 µL sur une cellule de Malassez (0.0624) et comptage de 5 carrés quadrillés
- Centrifuger 5 min à 80 g ; accélération 5, décélération 1
- Re-suspendre le culot dans une solution MMg pour avoir une concentration en protoplaste finale à 1.10^6 cellules/mL

Transfection :

- Mettre 20 µg de plasmide dans un tube eppendorf de 2 mL (fond rond) pour chaque réaction
- Ajouter 100 µL de solution de protoplastes et homogénéiser doucement par tapotements
- Ajouter goutte par goutte 110 µL de solution PEG et mélanger doucement après chaque goutte
- Incuber 10 min à température ambiante

- Ajouter 440 μL de solution W5 à température ambiante. Homogénéiser par tapotements
- Centrifuger 5 min à 80 g sans freinage
- Re-suspendre le culot dans 100 μL de solution W5
- Dans une plaque de culture cellulaire 24 puits : mettre 900 μL de milieu Gamborg complété + 100 μL de solution de protoplastes transfectés par puit
- Sceller la plaque avec du parafilm
- Incuber dans le noir, à 20°C avec agitation rotative 50 rpm