Genetic and epigenetic characterization of the geminivirus – host interaction

Tesis doctoral
Álvaro Piedra Aguilera

UNIVERSIDAD
DE MÁLAGA

AUTOR: Álvaro Piedra Aguilera

http://orcid.org/0000-0002-4208-5052

Esta Tesis Doctoral está depositada en el Repositorio Institucional de la Universidad de Málaga (RIUMA): riuma.uma.es
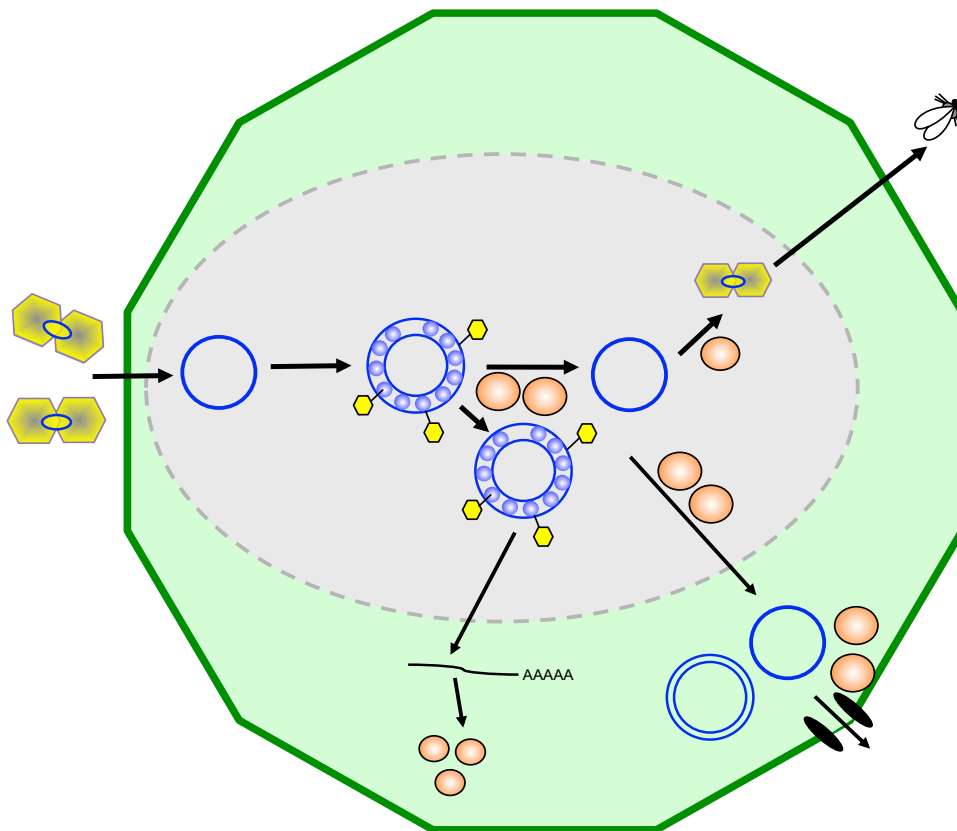
Genetic and epigenetic characterization of the geminivirus – host interaction

Tesis doctoral

Álvaro Piedra Aguilera

Programa de Doctorado Biotecnología Avanzada

Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora".
Universidad de Málaga – Consejo Superior de Investigaciones Científicas (UMA-CSIC).
Departamento de Biología Celular, Genética y Fisiología. Facultad de Ciencias.

Universidad de Málaga, 2019.

Directores de tesis:

- Dr. Eduardo Rodríguez Bejarano. Catedrático del Área de Genética del Departamento de
Biología Celular, Genética y Fisiología. Universidad de Málaga. IHSM-UMA-CSIC.
- Dra. Araceli Castillo Garriga. Profesora Titular del Área de Genética del Departamento Biología
Celular, Genética y Fisiología. Universidad de Málaga. IHSM-UMA-CSIC.

UNIVERSIDAD DE MÁLAGA

Dr. Eduardo Rodríguez Bejarano, Catedrático del Área de Genética del Departamento de Biología Celular, Genética y Fisiología de la Universidad de Málaga.

Dra. Araceli Castillo Garriga, Profesora Titular del Área de Genética del Departamento de Biología Celular, Genética y Fisiología de la Universidad de Málaga.

INFORMAN:

Que Don Álvaro Piedra Aguilera ha realizado bajo nuestra dirección y supervisión, en la Universidad de Málaga, el trabajo que bajo el título "Genetic and epigenetic characterization of the geminivirus-host interaction" presenta en esta memoria, la que constituye su tesis doctoral para aspirar al grado de Doctor en Biología.

Y para que así conste y tenga los efectos que correspondan, en cumplimiento de la legislación vigente, extienden el presente informe,

En Málaga, a 8 de Abril de 2019.

Fdo: Eduardo Rodríguez Bejarano   Fdo. Araceli Castillo Garriga

TRIBUNAL DE TESIS Y EVALUADORES EXTERNOS

Presidente
Dra. Carmen Beuzón López
Universidad de Málaga

Secretario
Dr. César Llave Correas
Centro de Investigaciones Biológicas (CIB-CSIC)

Vocal
Dr. Pedro Gómez López
Centro de Edafología y Biología Aplicada del Segura (CEBAS-CSIC)

Suplentes

Dr. Miguel Ángel Botella Mesa
Universidad de Málaga

Dra. Livia Donaire Segarra
Centro de Edafología y Biología Aplicada del Segura (CEBAS-CSIC)

Dra. Virginia Ruíz Ferrer
Universidad de Castilla-La Mancha

Evaluadores externos

Dr. Jesús Méndez Lozano
Instituto Politécnico Nacional, México

Dr. José Trinidad Ascencio Ibáñez
North Carolina State University

AGRADECIMIENTOS.

Si algo aprendí es que en ciencia no existe el "Yo", sino el "Nosotros". Poco o nada de este trabajo hubiera sido posible sin la inmensa ayuda que he recibido. ¡Muchas gracias desde ya!

En primer lugar a Araceli. Porque ha guiado mi investigación, hecho ver mis errores (que no han sido pocos) y ayudado a solucionarlos. Tantas tareas y problemas científicos (y administrativos también un montón) que encarar, hacia los cuales siempre puso su mejor esfuerzo. Tu ayuda fue imprescindible y siempre estaré agradecido.

A Eduardo. Por dejarme ser parte del laboratorio. Por participar muchas veces en lo experimental y en lo demás (por ser clave en conseguir la beca, entre otras cosas). Gracias por estar dispuesto a ayudar cada vez que lo necesité.

A Edgar. El que me enseñó gran parte de cómo se trabaja en un laboratorio. No puedo imaginarme una persona mejor con quien empezar. Menos mal que me tocó alguien con paciencia. Un buen amigo.

A los compañeros de Genética y Bioquímica, que para mí es lo mismo, y otros que hay por ahí. Me pregunto si en otros trabajos pasará que cuando tienes algo que resolver y pides ayuda al compañero/a, éste te responde con tan buena predisposición a solucionar tu problema. Porque eso es justo lo que me ha pasado con Miguel, Pepe, Blanca, Luis, Ana, Bea, Ángel V, Ale, Vítor, Noemí, Delphine, Bego, Lidia, Carmen, Lucía, Ali, Álvaro, Karen, Diego, Javi, Nieves, Jose R., David, Vito, Eva, Mario, Paco, Jessi, Jose V., Sonia, Manolo, Tábata, Natasa, Alba, Eloy, Adrián, Fran, Fernando, Jorge, Belén, Elena, Paqui, Almu, Jose A., Selene, Sara, Sito, Ángel, Miguel Ángel, Javi R. A., Carmen B., Victoriano, Juan Antonio, Aída, María, Eva M., Jorge S. y Carlos (Fundador y Presidente de C7 jeje). Si alguien piensa que debería estar en esta lista y no lo está, seguro que tiene razón, ¡fallo mío!

A Santi y los demás compañeros de fútbol. ¡Ya estoy de vuelta!

A mis padres y mi hermano. Más suerte no puedo tener. Lo bueno que pueda decir de ellos aquí se quedaría corto; además me da un poco de vergüenza. Ya sabéis lo que pienso de vosotros.

INDEX

Introduction

1.- *Geminiviridae* family.

1.1.- General characteristics.

Geminiviruses constitute a large family of plant viruses with circular single-stranded DNA (ssDNA) genome formed by one (monopartite) or two (bipartite) molecules (2.5-3.0 kb each one), which are packaged into two isometric particles fused to conform a geminate capsid (18 x 30 nm) (Figure 1A and 1B) (Zhang et al., 2001; Fondong, 2013). Geminivirus is one of the most devastating plant pathogens worldwide, causing very important crop losses such as tomatoes, cassava, grain legumes, maize, cotton, etc. Common symptoms associated with geminiviral infection include plant stunting, malformation of reproductive organs, curled and yellowing leaves, and sometimes, vein swelling and enations (Figure 1C) (Mansoor et al., 2003; Rojas et al., 2005; Briddon, 2009).
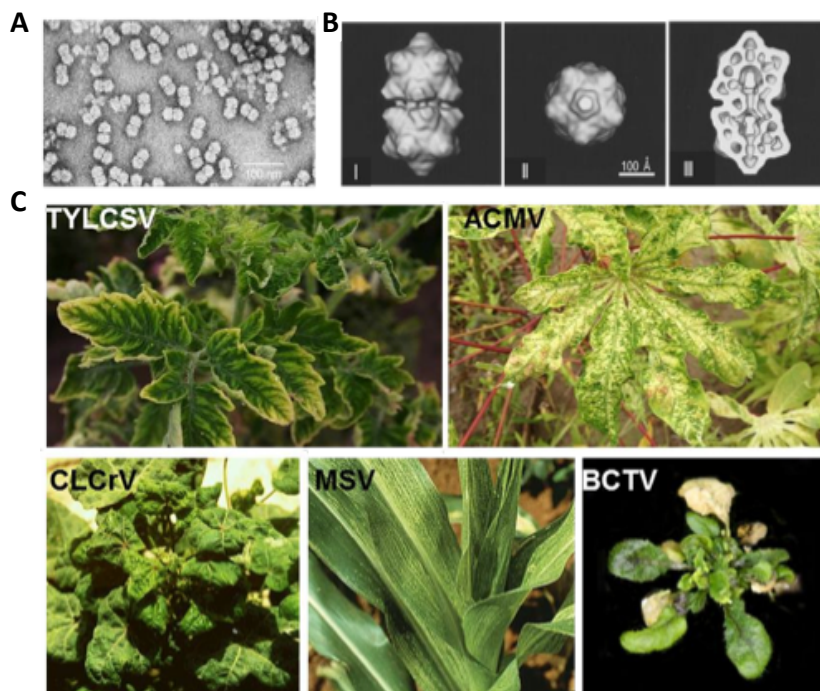
Figure 1. Geminiviral particles and symptoms. (A) Electron micrograph of geminivirus virion particles of *Maize streak virus* (MSV) (Zhang et al., 2001). **(B)** Reconstruction of MSV virion particle. Models were generated in silico from cryo-electron microscopy. Longitudinal (I) and transversal (II) views, and longitudinal section (III) (Zhang et al., 2001). **(C)** Symptoms of geminiviruses infection in different hosts: *Tomato yellow leaf curl Sardinia virus* (TYLCSV) in tomato, *African cassava mosaic virus* (ACMV) in cassava, *Cotton leaf crumple virus* (CLCrV) in cotton, *Maize streak virus* (MSV) in maize and *Beet curly top virus* (BCTV) in *Arabidopsis thaliana* (Mansoor et al., 2003; Rojas et al., 2005; Briddon, 2009).

1.2.- Geminivirus diversity.

The *Geminiviridae* family is classified into seven genera: *Becurtovirus, Begomovirus, Curtovirus, Eragovirus, Mastrevirus, Topocuvirus* and *Turncurtovirus,* according to the insect vector and the genome organization (Table 1) (Zerbini et al., 2017). The geminiviral genomes

encode six to eight open reading frames (ORFs), which are bi-directionally transcribed from the promoters located in the intergenic region (IR) (Rizvi et al., 2014). Among the *Geminiviridae* family, the *Begomovirus* and *Curtovirus* genera comprise most of the viral species infecting dicotyledonous plants.

Genus	Type member	Host range	Insect vector	Genome
Mastrevirus	*Maize streak virus* (MSV)	Monocots and dicots	Leafhopper (Fam. *Cicadelidae*)	Monopartite
Curtovirus	*Beet curly top virus* (BCTV)	Dicots	Leafhopper (Fam. *Cicadelidae*)	Monopartite
Begomovirus	*Bean golden mosaic virus* (BGMV)	Dicots	Whiteflies (*Bemisia tabaci*)	Mono- or bipartite
Topocuvirus	*Tomato pseudo-curly top virus* (TPCTV)	Dicots	Treehopper (Fam. *Membracidae*)	Monopartite
Becurtovirus	*Beet curly top Iran virus* (BCTIV)	Dicots	Leafhopper (*Circulifer haematoceps*)	Monopartite
Turncurtovirus	*Turnip curly top virus* (TCTV)	Dicots	Leafhopper (*Circulifer haematoceps*)	Monopartite
Eragrovirus	*Eragrotis curvula streak virus* (ECSV)	Monocots	Leafhopper (Fam. *Cicadelidae*)	Monopartite

Table 1. Recognized genera of the *Geminiviridae* family.

1.3.- Begomovirus and TYLCD: genome structure and proteins.

The *Begomovirus* genus consists of >320 species that infect dicots, are transmitted by the whitefly *Bemisia tabaci* and present monopartite or bipartite genomes (Zerbini et al., 2017). A large number of monopartite begomoviruses are associated with DNA satellites (Zhou, 2013).

The genome of monopartite begomoviruses is bi-directionally organized in two transcriptional units (Left, L and Right, R) separated by the intergenic region that contains key elements for initiating replication, including a stem-loop structure, and transcription of the viral genome (Hanley-Bowdoin et al., 1999; Fondong, 2013; Borah et al., 2016). The virion sense strand (VS strand) contains two ORFs in the transcriptional unit R, named V2 and V1, and the complementary sense strand (CS strand) encompasses four ORFs (C1, C2, C3 and C4) that are expressed from the transcriptional unit L (Figure 2A). V1 encodes the structural coat protein (CP), which in some monopartite begomoviruses also functions as viral DNA shuttle between the nucleus and the cytoplasm. V2 encodes a multi-functional protein involved in viral movement and the suppression of RNA-mediated gene silencing. Begomoviruses produce two proteins involved in viral replication: C1, encoding the replication initiator protein (Rep), which is essential for viral replication, and C3, encoding the replication enhancer protein (REn), which interacts with Rep and greatly enhances viral DNA accumulation in host cells. C4 is not critical for the virus infection but it is a pathogenicity determinant whose function has been associated to viral movement and RNA-silencing suppression. C2 is the transcriptional activator protein (TrAP) that activates the transcription of the genes in the R transcriptional unit (Jeske, 2009; Fondong, 2013; Borah et al., 2016).

Bipartite begomoviruses are composed of two DNA molecules of similar size (DNA-A and DNA-B) and the IRs from both molecules show a high sequence similarity. Both, DNA-A and DNA-B, are essential for the infection and are individually packaged. DNA-A is homologue to the monopartite genome. DNA-B contains two additional ORFs: BV1, encoding a movement protein (MP) that mediates viral DNA traffic cell to cell, and BC1, encoding a nuclear shuttle protein (NSP) involved in cytoplasm-nucleus movement (Figure 2B) (Fondong, 2013).
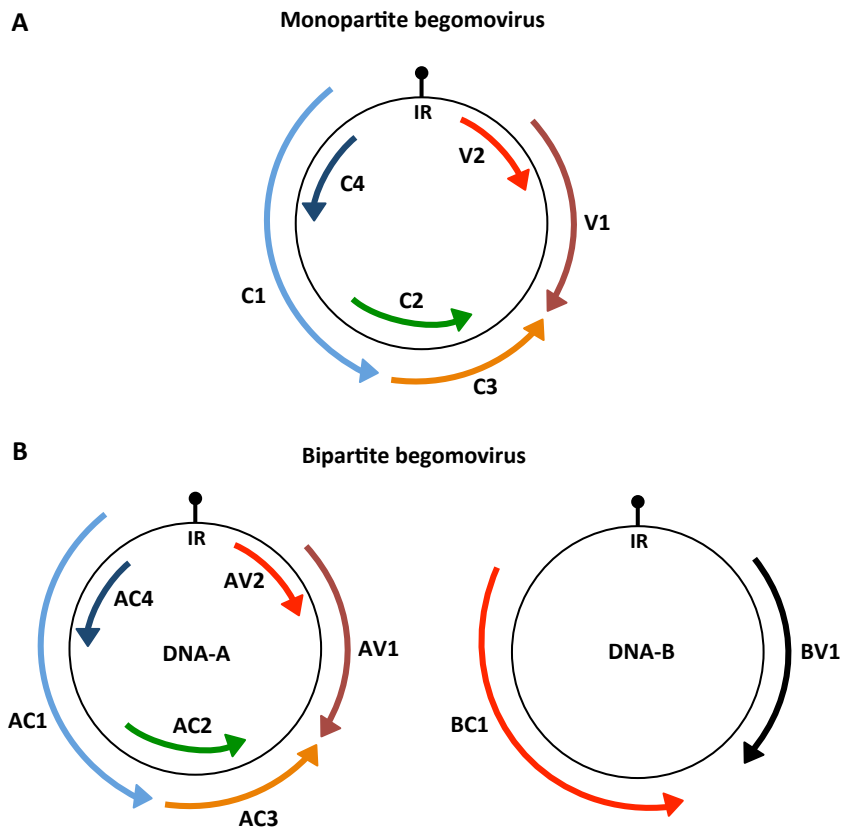
Figure 2. Genome organization of the *Begomovirus* genera. (A) Monopartite begomovirus. **(B)** Bipartite begomovirus.

Tomato yellow leaf curl disease (TYLCD), one of the most devastating viral diseases of cultivated tomato in tropical and subtropical regions worldwide (Díaz-Pendón et al., 2010; Navas-Castillo et al., 2011), is caused by the infection of several monopartite begomovirus species except for TYLCV from Thailand, which has two genomic components. The TYLCD complex comprises several species including: *Tomato yellow leaf curl Sardinia virus* (TYLCSV), *Tomato yellow leaf curl virus* (TYLCV), TYLCV-Mild (TYLCV-Mld) and *Tomato yellow leaf curl China virus* (TYLCCNV). Aggravated TYLCV disease symptoms in tomato are often associated with a TYLCV satellite DNAβ, as described in China (TYLCCNV DNAβ) (Zhou, 2013) and in Jordan (TYLCV DNAβ) (Anfoka et al., 2014). The symptoms of infected tomato plants are marginal chlorosis, upward curling of young leaves and drastic reduction in plant height because of the shortened internodes. TYLCV, the first known causal agent of TYLCD, was identified from severe outbreaks of the disease in the Jordan Valley (Cohen and Antignus, 1994). The disease has been reported in new areas since the middle 1990s, including the Mediterranean Basin, the Far East, America, and Australia. In Spain the disease was first observed in tomato plants in southern Spain in 1992 (Moriones et al., 1993). In the next years two more species were detected (TLCV and TYLC-Mld) (Navas-Castillo et al., 1997; Reina et al., 2000). Coinfection of plants with TYLCSV and TYLCV provoked the appearance of two recombinant viruses, *Tomato yellow leaf curl Malaga virus* (Monci et al., 2002) and *Tomato yellow leaf curl Axarquía virus* (García-Andrés et al., 2006).

1.4.- Geminiviral life cycle and replication.

After geminivirus acquisition by the insect vector, the single-stranded DNA (ssDNA, virion strand) is released from virions into the plant cell nucleus, and the complementary strand is synthesized by a plant DNA polymerase to yield double-stranded DNA (dsDNA). The dsDNA assembles with nucleosomes to form minichromosomes, and is transcribed by the plant RNA polymerase II. Once viral Rep is synthesized, it moves to the nucleus to initiate rolling-circle replication (RCR) and recombination-dependent replication (RDR) of the viral genome. REn interacts with Rep and enhances viral DNA accumulation. Geminiviral replication results in production of new ssDNA molecules. When the infection advances, Rep down-regulates its own expression and TrAP expression is activated, which promotes the activation of CP and NSP (NSP only in bipartite) expression. Then, the ssDNA can be packaged into CP and transported to the phloem, being available for the insect vector. On the other hand, CP (in monopartite) or NSP (in bipartite) can bind to and move viral DNA to the cytoplasm, where V2 (in monopartite) or MP (in bipartite) takes it through a plasmodesma to the next cell (Figure 3) (Rojas et al., 2001; Hanley-Bowdoin et al., 2013).
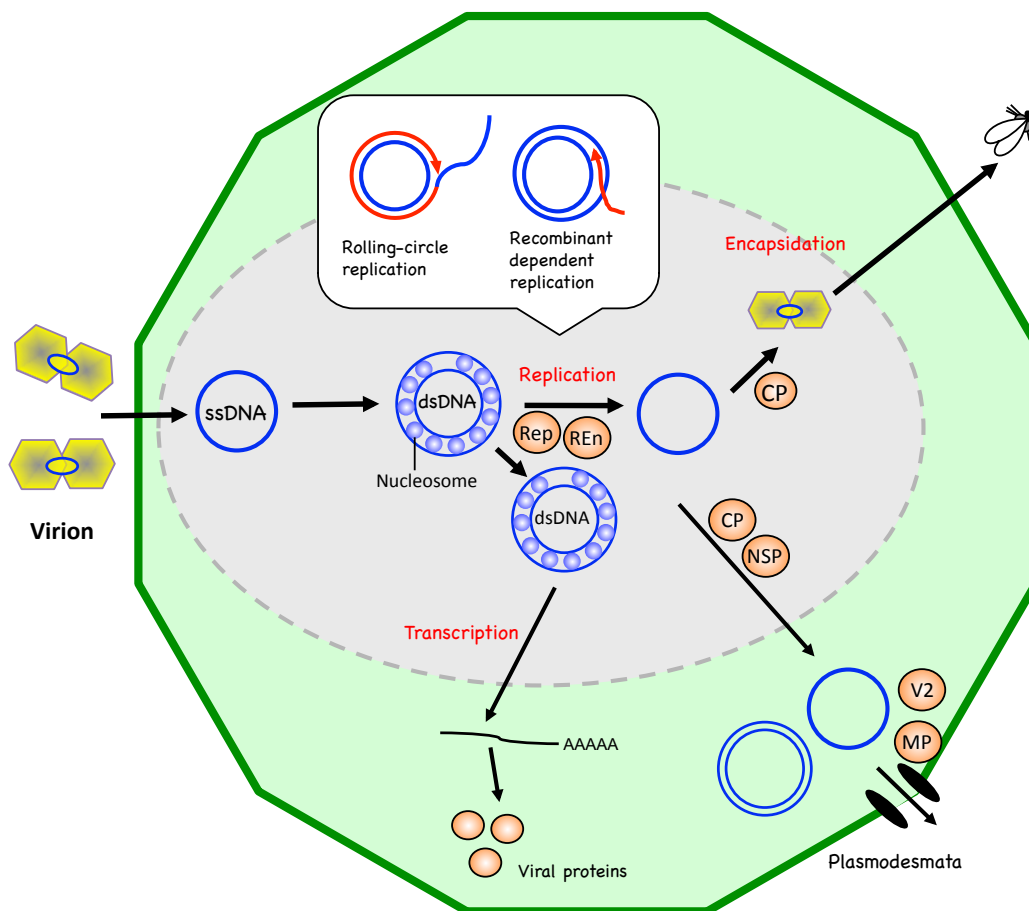
Figure 3. The begomovirus life cycle. Infection begins in a plant cell when geminiviral single-stranded DNA (ssDNA) is released from virions into the plant cell and copied into double-stranded DNA (dsDNA), which assembles with nucleosomes and is transcribed by the plant RNA polymerase II. Viral Rep protein initiates rolling-circle replication (RCR) and recombinant-dependent replication (RDR) to produce new ssDNA molecules. Later in infection, the ssDNA can be packaged into CP and transported to the phloem. On the other hand, CP or NSP can bind to and move viral DNA to the cytoplasm, where V2 or MP takes it through a plasmodesma to the next cell.

During rolling-circle replication (RCR), Rep nicks the virion strand of dsDNA molecules in a consensus nonanucleotide (TAATATTAC) sequence at the IR, and recruits the plant DNA replication machinery. The cleavage produced by Rep generates a 3'-end in the virion strand, which is extended by a DNA polymerase that uses the complementary strand as template. During the extension, the virion strand is displaced from the template strand, maintaining Rep linked to the 5'-end. Circular ssDNA is released by mean Rep-dependent cleavage and ligation of the virion strand. The ssDNA release occurs after one or more rounds of RCR, which produces one or more copies of ssDNA that can re-enter RCR (Figure 4A). An additional geminiviral replication mechanism is known as recombination-dependent replication (RDR). Viral linear ssDNA fragments invade a homologous sequence of the circular dsDNA and allow, by mean a plant DNA polymerase, the synthesis of linear dsDNA primed by the ssDNA fragments. Many copies of heterogeneous linear dsDNA are potentially produced and each copy could function as template for Rep synthesis. In turn, Rep triggers replication of linear dsDNA and the release of circular ssDNA that can enter RCR or RDR (Figure 4B) (reviewed in Jeske, 2009; Pooggin, 2013).
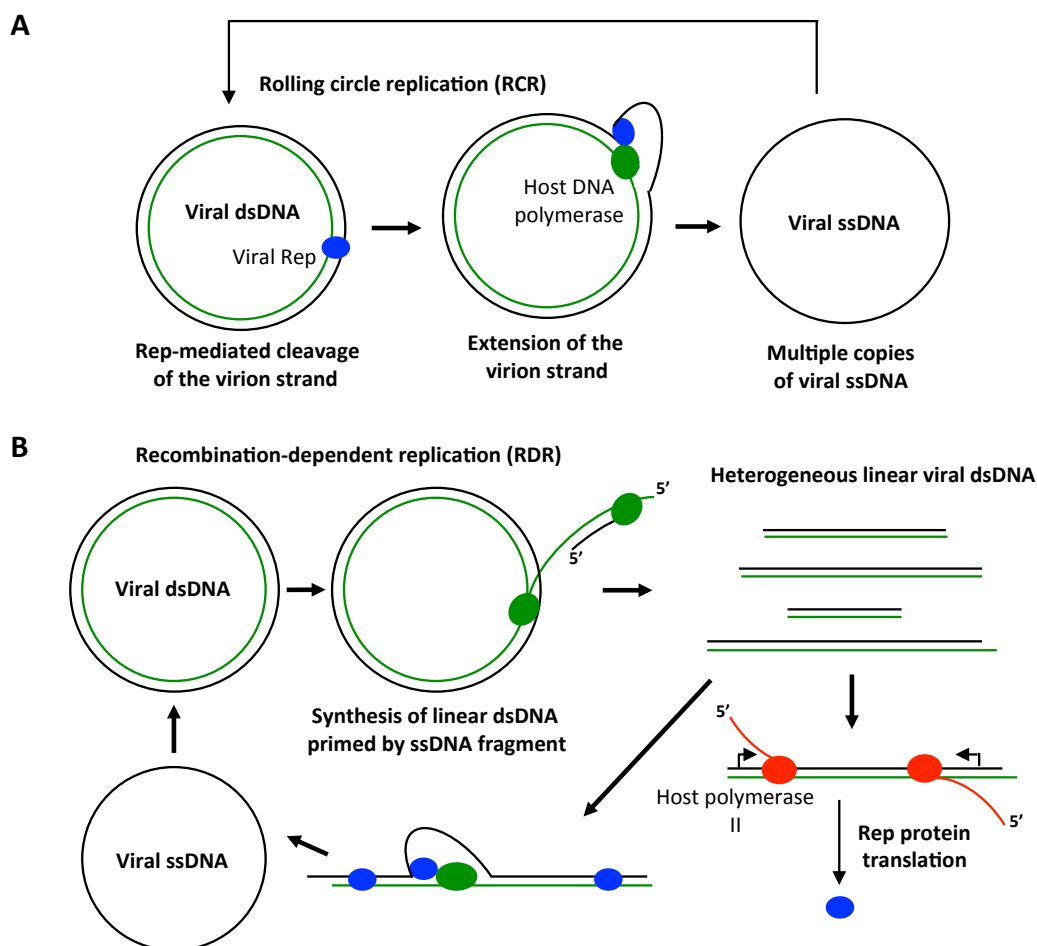
Figure 4. Models for RCR and RDR modes of geminivirus DNA replication. (A) Rolling-circle replication (RCR). Rep mediates the cleavage of the virion strand (black line), which is extended by a host DNA polymerase on the complementary strand (green line) template. After one or more rounds of replication, ore or more copies of ssDNA are released, which can re-enter RCR. **(B)** Recombination-dependent replication (RDR). Viral linear ssDNA fragments prime the synthesis of heterogeneous linear dsDNA, which is transcribed. Rep triggers the dsDNA replication and the release of ssDNA, which can enter RCR or RDR.

1.5.- Plant - geminivirus interaction.

Viruses reshape the intracellular environment during infection, both to co-opt processes necessary for the development of the viral infection and to interfere with antiviral defence. Studies of virus-host interactions have proven useful for understanding how viruses manipulate the host machinery and how re-wiring of specific pathways contribute to disease. Due to genome size constraints, geminivirus-encoded proteins are multifunctional regulators of plant cellular processes, such as replication, transcription and translation, which allow geminiviruses to establish a successful infection (Ascencio-Ibáñez et al., 2008; Pierce and Rey, 2013; Seo et al., 2018). Since geminiviruses do not encode their own DNA replication enzymes, their replication heavily relies on plant replication machinery. Besides, geminiviruses infect differentiated cells (vascular tissue or mesophyll cells) that have exited mitosis and display down-regulation of those proteins involved in DNA replication, and therefore, they require reprogramming the cell cycle to induce their own replication (Nagar et al., 1995; Egelkrout et al., 2001; Ascencio-Ibáñez et al., 2008). On the other hand, geminiviruses confront several plant defensive mechanisms involved in arresting viral replication and movement, such as: gene silencing, immune receptor signalling, hormone signalling, protein degradation and metabolism regulation (Rodríguez-Negrete et al., 2009; Hanley-Bowdoin et al., 2013; Calil and Fontes, 2016).

2.- Gene silencing mediated by RNA (RNA silencing).

RNA silencing refers to gene silencing processes that involve small RNAs (sRNAs). This mechanism plays a major role in the regulation of plant development and defence against biotic and abiotic stresses. It is triggered by double-stranded RNA (dsRNA) molecules, which can be produced as of transcription of self-complementary RNAs formed by inverted-repeated sequences, convergent transcription, or RNAs that are converted to dsRNA by mean RNA-dependent RNA polymerases (RDRs). An array of enzymes named ribonuclease III-type Dicer-like (DCL) proteins bind to and process the dsRNA into 20-25-nucleotides (nt) RNA duplexes that belong to two classes: small interfering RNAs (siRNAs) form populations coming from perfect or near perfect long dsRNAs, while the second class, miRNAs, are produced as discrete species arisen from imperfectly folded (hairpins structures) RNAs. RNA duplexes are stabilized by the methyltransferase HUA ENHANCER 1 (HEN1) that methylates the 3´-end of the sRNAs, and one strand is loaded into Argonaute (AGO) proteins. In turn, AGO binds to other proteins, forming the RNA-induced silencing complex (RISC). In the cytoplasm, RISC complex can target mRNAs complementary to loaded sRNAs to induce their post-transcriptional gene silencing (PTGS) through endonucleolytic cleavage or translational inhibition. In the nucleus, RISC (also named RNA-induced transcriptional silencing (RITS)) complex can induce transcriptional gene silencing (TGS) by directing repressive chromatin modifications (such as cytosine methylation and histone modifications) of *loci* that are complementary to loaded siRNAs (Pumplin and Voinnet, 2013; Matzke and Mosher, 2014).

2.1.- Post-transcriptional gene silencing (PTGS).

PTGS is mediated by mean miRNAs or siRNAs. In *A. thaliana*, miRNAs (21-nt, or less frequently, 22-nt) are commonly produced in the nucleus through DCL1 and its dsRNA-binding cofactor protein, HYPONASTIC LEAVES 1 (HYL1). In the cytoplasm, mature miRNAs are loaded into AGO1 to elicit PTGS. *A. thaliana* siRNAs can be loaded into AGO1 or AGO2 to mediate PTGS. 21- and 22-nt siRNAs are produced from dsRNA derived from non-coding RNA, mRNA or natural antisense RNA, by mean DCL4 and/or DCL2. A type of siRNA named trans-acting siRNA (tasiRNA) is formed when the single-stranded precursor is targeted and cleaved by the complex AGO-miRNA, which recruits RDR6 to convert the cleaved RNA fragment into dsRNA. Next, DCL4 processes dsRNA into a population of 21-nt tasiRNAs that are loaded into AGO1 and trigger PTGS. This phenomenon constitutes one of the mechanisms for RNA silencing amplification (Pumplin and Voinnet, 2013).

2.2.- Transcriptional gene silencing (TGS): RNA-directed DNA methylation (RdDM) and its interplay with histone modifications.

DNA methylation is a reversible epigenetic mark that regulates gene expression, genetic imprinting and silencing of transposable elements and repeats, controlling the compacted state of the chromatin (Gehring, 2013, Baulcombe and Dean, 2014). DNA methylation constitutes a defence against invasive nucleic acid such as transposons, transgenes and viruses. In plants, methylation occurs at cytosines in symmetric (CG or CHG) or asymmetric (CHH) sequence contexts (H: C, T or A) (Matzke and Mosher, 2014).

RNA-directed DNA methylation (RdDM) is a pathway that regulates gene expression by eliciting DNA methylation at sites lacking previously this mark; thus, it is also named *de novo* DNA methylation. Two RNA polymerases participate in this pathway, DNA-DEPENDENT RNA POLYMERASE IV (Pol IV) and Pol V, and both are plant-specific enzymes evolved from Pol II. The recruitment of Pol IV to target sequences, which are primarily transposons and other repeats, is not fully understood. Pol IV is recruited to a subset of its genomic targets by the Pol IV-interacting protein SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1), which binds to the methylated lysine 9 of histone H3 (H3K9me) and unmethylated lysine 4 of histone H3 (H3K4). Pol IV transcripts are copied into long dsRNAs by RDR2, and the dsRNAs are processed by DCL3 into 24-nt siRNAs that are loaded into AGO4. Similarly to Pol IV, the chromatin features responsible for recruiting Pol V to its target sequences remain incompletely understood. The nascent transcripts generated by Pol V are targeted by the AGO4-24-nt siRNA complex by sequence complementarity. This targeting recruits the *de novo* methyltransferases DOMAINS REARRANGED METHYLTRANSFERASE 1 and 2 (DRM1 and DRM2), which leads to cytosine methylation at the three contexts and TGS. Other proteins that mediate Pol V transcription and DRM2 binding are DEFECTIVE IN RNA-DIRECTED DNA METHYLATION (DRD1), DEFECTIVE IN MERISTEM SILENCING 3 (DMS3) and RNA-DIRECTED DNA METHYLATION 1 (RDM1) (Figure 5)

(reviewed in Matzke and Mosher, 2014). Pol V could also be recruited to some target sequences by three histone methyltransferases that bind to methylated DNA, SUVH2, SUVH9 and SUVR2, which are members of the SU(VAR)3-9 histone methyltransferase family. This methyltransferase family is responsible for the methylation of histone H3 at lysine 9 (H3K9me), a hallmark for heterochromatin. The interplay between RdDM and histone methylation at lysine 9 reinforces TGS. Around 70% of RdDM targets contain H3K9me, a repressive histone mark catalysed by the SUVH histone methyltransferases SUVH4 (named KRYPTONITE, KYP), SUVH5 and SUVH6. Additionally, the removal of activation histone marks, such as H3K4me, allows that some RdDM targets maintain DNA methylation (Matzke and Mosher, 2014; Pikaard and Mittelsten, 2014).
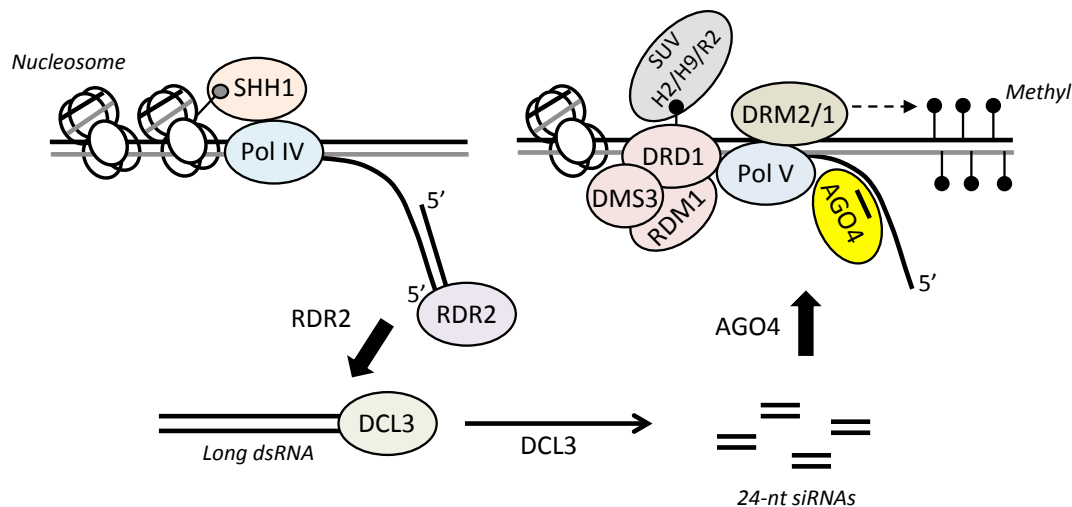
Figure 5. Model for RNA-directed DNA methylation (RdDM) (*de novo* methylation). Pol IV could be recruited to genomic targets by SHH1, which binds to H3K9me. RDR2 copies the Pol IV transcripts into long dsRNA, which is processed by DCL3 into 24-nt siRNAs that are loaded into AGO4. By sequence complementarity, the Pol V transcripts are targeted by the AGO4-24-nt siRNA complex. This targeting recruits two proteins, DRM1 and DRM2, involved in the DNA methylation at the three cytosine contexts. DRD1, DMS3 and RDM1 mediate Pol V transcription and DRM2 binding. SUVH2, SUVH4 and SUVR2 bind to methylated DNA and recruit Pol V to some target sequences. Black lollypops represent DNA methylation at CG, CHG or CHH. Grey lollypop represents H3K9me.

A complex set of maintenance mechanisms ensures the persistence of established DNA methylation following DNA replication. MET1 (METHYLTRANSFERASE 1) efficiently maintains CG methylation aided by VARIANT IN METHYLATION 1 (VIM1), VIM2 and VIM3, three proteins that bind to CG-specific methyl, while CMT2 (CHROMOMETHYLASE2) and CMT3 maintain methylation at CHH and CHG sites, respectively (Stroud a et al., 2013). Non-CG context methylation (CHG and CHH) is also maintained by a self-reinforcing loop involving KYP (KRYPTONITE) family enzymes that recognize methylated DNA (non-CG context) and methylate the lysine 9 of neighbouring histone H3 (H3K9me2). In newly replicated DNA, CMT2 and CMT3 bind to H3K9me2 and maintain the methylation of adjacent CHH and CHG sites, respectively (Figure 6). CHH methylation is also maintained through persistent *de novo* methylation by DRM1/2 through RdDM (Law and Jacobsen 2010; Du et al., 2012; Stroud et al., 2013). The access to DNA of MET1, CMT2 and CMT3 is facilitated by DECREASED DNA METHYLATION 1 (DDM1), a nucleosome remodeler that

hydrolyses ATP to move along DNA and alter nucleosome composition and placement (Figure 6) (Zemach et al., 2013).
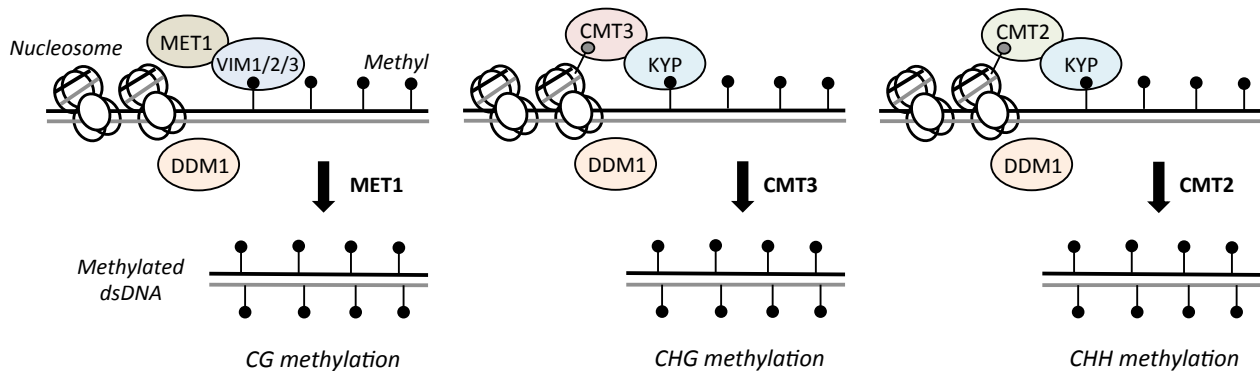
Figure 6. Models for maintenance DNA methylation at CG, CHG or CHH. After DNA replication, the newly synthesized DNA strand (grey line) is methylated by MET1 (CG methylation, aided by VIM1/2/3), CMT3 (CHG methylation, aided by KYP) or CMT2 (CHH methylation, aided by KYP). The access to DNA of MET1, CMT3 and CMT2 is facilitated by DDM1. Black lollypops represent DNA methylation at CG, CHG or CHH. Grey lollypops represent H3K9me2.

DNA demethylation plays an important role in the regulation of the gene expression. The demethylation process can be passive, by mean the absence of maintenance methylation following DNA replication, or active, through the action of DNA demethylases. The characterization of the DNA demethylases ROS1 (REPRESSOR OF SILENCING 1), DME (DEMETER), DML2 (DEMETER-LIKE 2) and DML3 (DEMETER-LIKE 3) in *A. thaliana* suggests that subsets of genomic DNA methylation patterns are the products of antagonistic methylation-demethylation activity (Gong et al., 2002; Penterman et al., 2007). Both DME and ROS1 catalyse the release of 5-meC from DNA by a glycosylase/lyase mechanism (Morales-Ruíz et al. 2006), not only at CG but also non-CG sequences. DML2 and DML3 are ROS1 and DME paralogs expressed in a wide range of plant organs and are also required for maintenance of proper DNA methylation patterns (Morales-Ruíz et al. 2006; Gehring et al. 2006, Ortega-Galisteo et al., 2008). The triple demethylase mutant *ros1 dml2 dml3* (DME expression is just detected in the female gametophyte) causes increased DNA methylation in all sequence contexts, not genome-wide but at specific genomic *loci* (Lister et al., 2008).

2.3.- RNA silencing as a plant defence mechanism against viruses.

RNA silencing is an important antiviral defence mechanism in plants (Pumplim and Voinnet 2013). Viral double-stranded RNA (dsRNA) is recognized by a set of DCL ribonucleases and processed into 21 to 24-nt primary viral small RNAs (vsRNAs). In *A. thaliana,* DCL4, which produces 21-nt vsRNAs, is the primary sensor of viral dsRNAs; DCL2 acts as a DCL4 surrogate, producing 22-nt vsRNAs, while DCL3 targets dsRNA to generate 24-nt vsRNAs (Donaire et al., 2009; Díaz-Pendón et al., 2007). DCL1 is only a minor contributor to vsRNAs formation for plant RNA and ssDNA viruses (Moissiard and Voinnet, 2006). In DNA viruses, the generation of dsRNA

molecules might derive from overlapping of convergent transcription units, as part of aberrant viral transcripts, or from viral sequences with internal stem-loop structures/hairpin structures (Chellappan et al., 2004; Donaire et al., 2009; Aregger et al., 2012). In a RNA silencing amplification step, one or more host-encoded RDRs use viral ssRNA to synthesize dsRNA, which serves as a substrate for the generation of secondary vsRNAs by DCL proteins. In *A. thaliana,* RDR1, RDR6 and possibly RDR2 have been implicated in vsRNA production (Pumplin and Voinnet, 2013). Both primary and secondary vsRNAs support the systemic silencing that spreads throughout the plant. VsRNAs associate with distinct AGO-containing complexes, which provide targeting specificity for RNA degradation or translational repression (PTGS), or chromatin modification (DNA methylation and histone modification, TGS) through a sequence homology-dependent mechanism (Bologna and Voinnet, 2014; Borges and Martienssen, 2015).

3.- Geminivirus and transcriptional gene silencing (TGS).

3.1.- Interaction of geminivirus and the plant DNA and histone methylation machinery.

Geminiviral dsDNA associates with plant histones to form minichromosomes (Pilartz and Jeske, 1992; 2003), which carry the same epigenetics marks as the plant chromatin, i.e. cytosine and histone methylation (Raja et al., 2008, Castillo-González et al., 2015, Ceniceros-Ojeda et al., 2016). Reported levels of geminiviral DNA methylation in infected plants depend on the geminivirus-plant system, ranging from 50% to 2% of viral cytosines methylated (Bian et al., 2006; Raja et al., 2008; Rodríguez-Negrete et al., 2009; Yadav and Chattopadhyay 2011; Yang et al., 2011b; Zhang et al., 2011; Paprotka et al., 2011; Sahu et al., 2014; Jackel et al., 2016; Deuschle et al., 2016; Torchetti et al., 2016; Ceniceros-Ojeda et al., 2016). Most of the reports on geminivirus DNA methylation base their data on sequencing only a few clones (from 4 to 20) generally from the IR and surrounding sequences, which are PCR-amplified after bisulfite treatment. This approach gives a very limited picture of the exact geminivirus DNA methylation pattern. In general, low levels of DNA methylation are reported for geminivirus in sensitive plants, which has led to the hypothesis that geminiviruses evade repressive cytosine methylation and TGS very efficiently by their resourceful replication mechanisms (Paprotka et al., 2011; Pooggin, 2013). On the other hand, in tissues showing host recovery (emergence of asymptomatic leaves after a typical symptomatic infection), an increase of the viral DNA methylation levels or minichromosome compaction is detected, supporting an active role of TGS in host defence. Raja et al. (2008) detected that *A. thaliana* plants recovered from the infection with the curtovirus *Beet curly top virus* (BCTV) lacking a functional C2 (BCTV C2-), but not from a wild-type BCTV infection. Such recovery resulted methylation-dependent since AGO4 protein was required. Furthermore, viral DNA from the IR of the mutant virus obtained from recovered secondary tissue resulted hypermethylated as compared with primary symptomatic tissue. Recently, Coursey and collaborators (2018) demonstrated that the BCTV C2- mutant virus presented higher viral chromatin compaction and nucleosome occupancy in *A. thaliana*

recovered tissue, compared to symptomatic tissue. This correlated with increased methylation levels of viral DNA and H3K9, and reduced virus amount, being Pol IV- and Pol V-dependent (Jackel et al., 2016). On the other hand, by using *Pepper golden mosaic virus* (PepGMV)-infected pepper plants, Rodríguez-Negrete and collaborators (2009) observed a recovery phenotype associated with a reduction of viral accumulation and an increase of methylation levels of the viral IR. A more recent work using the same pathosystem revealed that geminiviral minichromosomes isolated from symptomatic tissue were associated to chromatin activation marks (H3K4me3) and low DNA methylation levels. Conversely, minichromosomes isolated from recovered tissue harboured chromatin repressive marks (H3K9me2) and high DNA methylation levels (Ceniceros-Ojeda et al., 2016).

A direct evidence for DNA and histone methylation of viral minichromosomes as an antiviral defence mechanism arose from experiments showing that *A. thaliana* and *N. benthamiana* plants impaired in different components of the maintenance, *de novo* DNA methylation or histone methylation pathways, displayed enhanced susceptibility to geminivirus infection (Raja et al., 2008; Sun et al., 2015; Castillo-González et al., 2015). Such increased susceptibility correlated with reduced viral DNA methylation (Raja et al., 2008).

3.2.- Geminiviral TGS suppressors.

Several geminiviral proteins (C2, Rep and V2, and the geminivirus-associated betasatellite βC1) have been shown to suppress TGS (Hanley-Bowdoin et al., 2013; Wang et al., 2014; Castillo-González et al., 2015). The best-characterized TGS suppressor is C2, which interferes with different enzymes of the plant methyl cycle. C2 from the bipartite begomovirus *Tomato golden mosaic virus* (TGMV) and the curtovirus BCTV were observed to inactivate ADENOSINE KINASE (ADK), an enzyme involved in the synthesis of the essential methyltransferase cofactor S-adenosyl methionine (SAM) (Wang et al., 2003). Buchmann and collaborators (2009) observed that ectopic expression of TGMV and BCTV C2 in *N. benthamiana* inhibited the TGS of a silenced GFP transgene. This occurred through a mechanism that does not depend on transcriptional activation, pointing to the impairment of the methyl cycle. C2 from the bipartite begomovirus *Cabbage leaf curl virus* (CaLCuV) reduced the DNA methylation levels and increased the expression of TGS-silenced *A. thaliana* loci (Buchmann et al., 2009). Together, these results suggest that C2 reverts TGS by inhibiting ADK. Additionally, Zhang and collaborators (2011) identified that C2 from the curtovirus *Beet severe curly top virus* (BSCTV) interacts with an *A. thaliana* S-adenosyl-methionine decarboxylase, SAMDC1, and inhibits its proteasome-mediated degradation. SAMDC1 catalyzes the conversion of SAM to the descarboxylated form (dcSAM), thus reducing the accumulation of this methyltransferase cofactor. On the other hand, a work from Castillo-González et al. (2015) pointed to the TGS suppressor activity of TGMV and CaLCuV C2 at the macrostructural level. They described that TGMV and CaLCuV C2 bound to and inhibited the catalytic activity of KYP *in vitro*, while *in vivo* both proteins reduced host H3K9me2 levels. It was observed that KYP bound to viral minichromosomes, which harboured H3K9me2;

importantly, *kyp* null mutants deployed a decrease of H3K9me2 marks on several viral *loci* (Castillo-González et al., 2015). These results highlight the role of C2 as TGS suppressor by mean inhibition of KYP.

The geminiviral Rep protein was also described as TGS suppressor. Rodríguez-Negrete et al. (2013) observed that several geminiviruses repressed the expression of the maintenance DNA methyltransferases MET1 and CMT3 (but not the *de novo* DNA methyltransferase DRM2) in *N. benthamiana* and *A. thaliana*. They also demonstrated the down-regulation of these genes depends on Rep. In *A. thaliana*, they showed the *MET1* and *CMT3* down-regulation upon TYLCSV Rep expression correlated with increased transcript levels of TGS-silenced *loci* that are predominantly controlled by CG methylation, and reduced cytosine methylation at CG sites was found within those reactivated *loci*. Additionally, they described that the ectopic expression of TYLCSV Rep reverted the silencing of a TGS-controlled GUS transgene in *A. thaliana*.

The third known TGS suppressor is V2 from TYLCV. Wang et al. (2014) observed that V2 reverted the TGS of a transcriptionally silenced GFP transgene in *N. benthamiana*. Furthermore, upon V2 ectopic expression, they detected a reduction of methylation levels in the methylation-controlled *A. thaliana loci AtSN1* and *AtMEA-ISR* (Wang et al., 2014).

Finally, DNA molecules related to geminivirus have been characterized to revert TGS. Some geminiviruses associate with other DNA molecules named betasatellites (DNA β), whose replication depends on the helper virus. In turn, the protein encoded by the satellite enhances viral replication and triggers symptoms. The βC1 protein from the betasatellite TYLCCNB, associated with the geminivirus TYLCCNV, reduced the methylation levels of the virus and the plant genome. Moreover, the silencing of a GFP transgene and an endogenous *locus* was reverted upon TYLCCNB infection (Yang et al., 2011b). Yang and collaborators linked these findings to the interaction between βC1 and S-ADENOSYL HOMOCYSTEINE HYDROLASE (SAHH), a different enzyme that participates in the methyl cycle and is involved in SAM production.

Materials and methods

1.- Microorganisms: bacterial strains and growth conditions.

Manipulation of bacteria was carried out according to standard methods (Sambrook and Russell, 2001). *Escherichia coli* and *Agrobacterium tumefaciens* bacteria were used in this work.

1.1.- *Escherichia coli*.

Escherichia coli DH5α strain (genotype: *F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG φ80δλαχΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ-*) was used for cloning procedures. It was grown in Luria-Bertani (LB) medium (Maloy, 1990) supplemented with ampicillin (100 µg/mL) or kanamycin (50 µg/mL) for plasmid selection, at 37ºC.

1.2.- *Agrobacterium tumefaciens*.

Three *Agrobacterium tumefaciens* strains were used in this work: GV3101, C58c1 and LBA4404. GV3101 and C58c1 strains were used for agroinoculation assays of *Arabidopsis thaliana* and *Nicotiana benthamiana*, whereas LBA4404 strain was used for agroinoculation of *Solanum lycopersicum* cv. Moneymaker. Bacteria were grown in LB medium supplemented with the corresponding antibiotics (kanamycin (50 µg/mL), tetracycline (10 µg/mL), rifampicin (50 µg/mL)), at 28ºC.

2.- Constructs and primers.

2.1.- Constructs generated in this work.

Previously, Dr. Rodríguez-Negrete obtained by RT-PCR, partial cDNA clones for *NbMET1* (500 bp), *NbCMT3* (591 bp), *NbDRM2* (567 bp) and *NbROS1* (372 bp). PCR products were cloned into the pGEM-T easy vector (Promega) to yield pGENbMET1, pGENbCMT3, pGENbDRM2 and pGENbROS1 (Rodríguez-Negrete et al., 2013). Constructs were fully sequenced, and partial cDNA sequences were submitted to GenBank database (AC: *NbMET1*, FJ222441; *NbCMT3*, JQ957858; *NbDRM2*, JQ957857; *NbROS1*, JQ957859). In this work, the fragment corresponding to the partial cDNA was digested with the *Apa*I and *Spe*I restriction enzymes (Takara), and the digestion products were run in a 1% EtBr-stained agarose gel for visualization. The gel band corresponding to the *NbMET1*, *NbCMT3, NbDRM2* or *NbROS1* fragment was purified using GFX PCR DNA and Gel Band Purification kit (GE Healthcare), according to the manufacturer's instructions. Next, fragments were subcloned into *Apa*I/*Spe*I sites of *Tobacco rattle virus* (TRV) RNA2-based pTV00 vector (Ratcliff et al., 2001) to yield the TRV-based constructs (Figure 1).
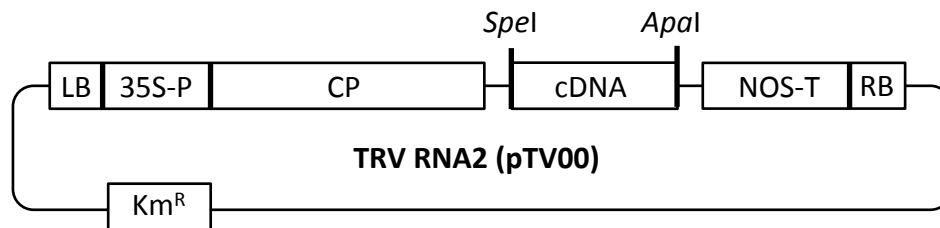
Figure 1. Genomic organization of TRV-based constructs. Partial cDNA fragments of *NbMET1*, *NbCMT3, NbDRM2* and *NbROS1* were cloned into *Apa*I/*Spe*I sites of *Tobacco rattle virus* (TRV) RNA2-based pTV00 vector (Ratcliff et al., 2001). The cDNA fragments are positioned between the left and right border (LB and RB), and between CaMV 35S promoter (35S-P), Nopaline synthase terminator (NOS-T) and the coat protein (CP) of TRV. Backbone contains a kanamycin resistance (Km^R) cassette.

2.2.- Constructs used in this work.

Constructs used in this thesis are listed in Table 1.

Table 1. Constructs used in this thesis.

Plasmid	Vector backbone	Fragment	Strain	Used for	Collection number	Reference
pGENbMET1	pGEM-T easy	*NbMET1*	DH5α	Cloning	3908	Dr. Rodríguez-Negrete, unpublished
pGENbCMT3	pGEM-T easy	*NbCMT3*	DH5α	Cloning	4806	Rodríguez-Negrete et al., 2013
pGENbDRM2	pGEM-T easy	*NbDRM2*	DH5α	Cloning	4808	Rodríguez-Negrete et al., 2013
pGENbROS1	pGEM-T easy	*NbROS1*	DH5α	Cloning	4812	Rodríguez-Negrete et al., 2013
pTVNbMET1	pTV00	*NbMET1*	DH5α	Cloning	3963	Dr. Rodríguez-Negrete, unpublished
pTVNbCMT3	pTV00	*NbCMT3*	DH5α	Cloning	5532	This work
pTVNbDRM2	pTV00	*NbDRM2*	DH5α	Cloning	5550	This work
pTVNbROS1	pTV00	*NbROS1*	DH5α	Cloning	5535	This work
pGTYS	pGreen-0229	TYLCSV genome	GV3101	Plant infection	2419	Lozano-Durán et al., 2011
pBINTYM	pBIN19	TYLCV genome	GV3101	Plant infection	5568	Navas-Castillo et al., 1999
pBINBC	pBINX1	BCTV genome	GV3101	Plant infection	2543	Briddon et al., 1989
pGCaLA	pGreen-0229	CaLCuV DNA A genome	GV3101	Plant infection	6262	José Cana Quijada, unpublished
pGCaLB	pGreen-0229	CaLCuV DNA B genome	GV3101	Plant infection	6264	José Cana Quijada, unpublished
pBINIL	pBINX	TYLCV-IL genome	LBA4404	Plant infection		Supplied by Dr. Díaz-Pendón
pBINX1	pBINX1	Empty	GV3101	Plant infection (mock control)	2375	Sánchez-Durán et al., 2011
pTVNbMET1	pTV00	*NbMET1*	GV3101	VIGS	3965	This work
pTVNbCMT3	pTV00	*NbCMT3*	GV3101	VIGS	5542	This work
pTVNbDRM2	pTV00	*NbDRM2*	GV3101	VIGS	5564	This work
pTVNbROS1	pTV00	*NbROS1*	GV3101	VIGS	5546	This work
pTVNbPDS	pTV00	*NbPDS*	GV3101	VIGS	1456	Ratcliff et al., 2001
pTV00	pTV00	TRV2	GV3101	VIGS (mock control)	1455	Ratcliff et al., 2001
pBINTRA6	pBINTRA6	TRV1	C58c1	VIGS	1457	Ratcliff et al., 2001

2.3.- Primers used in this work.

Primers used in this thesis are listed in Table 2.

Table 2. Primers used in this thesis. Asterisks indicate primers used for sqPCR.

qPCR / sqPCR				
Primer name	**Primer sequence (5' - 3')**	**Target gene (template: DNA / cDNA)**	**Reference**	**Collection number**
ACT2-F	CTAAGCTCTCAAGATCAAAGGCTTA	*A. thaliana Actin* (DNA)	Rodríguez-Negrete et al., 2013	902
ACT2-R	ACTAAAACGCAAAACGAAAGCGGTT			903
NbACT-F	TCACAGAAGCTCCTCCTAATCC	*N. benthamiana Actin* (DNA)	Dr. Rodríguez-Negrete, unpublished	1375
NbACT-R	GGGAAGAACAGCCTGAATG			1376
EF1α-F	GATTGGTGGTATTGGAACTGTC	*N. benthamiana EF1a* (DNA and cDNA)	Rotenberg et al., 2006	1379
EF1α-R	AGCTTCGTGGTGCATCTC			1380
25S-F	ATAACCGCATCAGGTCTCCA	*N. benthamiana 25S* (DNA and cDNA)	Mason et al., 2008	1192
25S-R	CCGAAGTTACGGATCCATTT			1193
C4TYM-F	AAGCAGGGCAGCACATTTCCATC	TYLCV Rep/C4 (DNA)	Rodríguez-Negrete et al., 2013	1485
C4TYM-R	CTGCGGCGTAAGCGTCATTG			1486
C4BC-F	CTACACGAAGATGGGCAACCT	BCTV Rep/C4 (DNA)	Rodríguez-Negrete et al., 2013	1487
C4BC-R	TGACGTCGGAGCTGGTTTAG			1488
C4TS-F	TCCCCAACCAGATCAGCACAT	TYLCSV Rep/C4 (DNA)	Rodríguez-Negrete et al., 2013	1483
C4TS-R	TTGGCGTAAGCGTCATTGG			1484
C4CaLCuV-F	AATATGTCGGCCCAGAAGC	CaLCuV Rep/C4 (DNA)	Mauricio Zúñiga Marín, unpublished	1523
C4CaLCuV-R	CAGGATACAACGCGTAGAGCA			1524
VT-NbMET1-F	GCTCCATTCCGTTCGCTTACTGT	*N. benthamiana MET1* (cDNA)	Dr. Rodríguez-Negrete, unpublished	1788
VT-NbMET1-R	ATTCTTTGGCACCTGATTAG			1789
VT-NbCMT3-F	GAACCCCATAATCAGACCATAATACA	*N. benthamiana CMT3* (cDNA)	Dr. Rodríguez-Negrete, unpublished	1784
VT-NbCMT3-R	TTCATCGCCATCGCCAAAGA			1785
VT-NbROS1-F	CCAAATAGCTACGAGGACAGACA	*N. benthamiana ROS1* (cDNA)	Dr. Rodríguez-Negrete, unpublished	1479
VT-NbROS1-R	TCGTTTTTGAAGTTACTACAGATGG			1480
TRV-F	GTGCACGCAACAGTTCTAATCG	TRV (cDNA)	Fernández-Calvino et al., 2014	1495
TRV-R	GCTGTGCTTTGATTTCTCCACC			1496
25S2-F	TTTGGTAAGCAGAACTGGCG	*N. benthamiana 25S* (DNA and cDNA)	This work	2148
25S2-R	CGTCCTGCTGTCTTAATCGACC			2149
18S-F	TTCCGTTAACGAACGAGACC	*N. benthamiana 18S* (DNA and cDNA)	This work	2156
18S-R	TGTCGGCCAAGGCTATAAAC			2158
5.8S-F	CGCATCGATGAAGAACGTAG	*N. benthamiana 5.8S* (DNA and cDNA)	This work	2153
5.8S-R	CCTCGGCCTAATGGCTTC			2155
5S-F	GCGATCATACCAGCACTAAGC	*N. benthamiana 5S* (DNA)	This work	2150
5S-R	GGATGCAACACGAGGACTTC			2152
Rubisco-F*	TACTTGAACGCTACTGCAG	*N. benthamiana Rubisco* (cDNA)	Castellano et al., 2015	
Rubisco-R*	CTGCATGCATTGCACGGTG			
25ST-F*	CACCAATAGGGAACGTGAGCTG	*N. benthamiana 25S* (cDNA)	Castellano et al., 2015	
25ST-R*	CCTCTTTTCGGAAAAGTTAGAAATTG			
At25S-F	ATAACCGCATCAGGTCTCCAAG	*A. thaliana 25S* (DNA)	Iskandar et al., 2004	
At25S-R	CCTCAGAGCCAATCCTTTTCC			
At18S3-F	TTCCGTTAACGAACGAGACC	*A. thaliana 18S* (DNA)	This work	
At18S3-R	TGCGGCCCAGAACATCTAAG			

T4 DNA polymerase DNA synthesis				
Primer name	**Primer sequence (5' - 3')**	**Description**	**Reference**	**Collection number**
TAG-370(+)	AGTTTAAGAACCCTTCCCGCGAAGGCTGAACTTCGACAGC	DNA synthesis of CS strand	Rodríguez-Negrete et al., 2014	866
TAG-540(-)	AGTTTAAGAACCCTTCCCGCGGACTTTACATGGGCCTTCAC	DNA synthesis of VS strand	Rodríguez-Negrete et al., 2014	867

TYLCV-IL VS, CS and total DNA quantification				
Primer name	**Primer sequence (5' - 3')**	**Description**	**Reference**	**Collection number**
TYLCD CP-F	GAAGGCTGAACTTCGACAGC	VS (+TAG) and total DNA (+TYLCD CP-R) quantification	Rodríguez-Negrete et al., 2014	1473
TYLCD CP-R	GGACTTTACATGGGCCTTCAC	CS (+TAG) and total DNA (+TYLCD CP-F) quantification	Rodríguez-Negrete et al., 2014	1474
TAG	AGTTTAAGAACCCTTCCCGC	VS and CS quantification	Rodríguez-Negrete et al., 2014	865

Genotyping				
Primer name	Primer sequence (5' - 3')	Target gene (template: DNA)	Reference	Collection number
METNF2	TAGCCAACAAGTTATCGCTTACT	*A. thaliana MET1*	Johnson et al., 2007	2003
METNR2	TTCGCAAACCATTCTTCACAGAG			2004
TL-4	TAATTGCGTCGAATCTCAGCATC	T-DNA insertion	Johnson et al., 2007	2005
drm1-LP	GAGCCGTCTCATCAAACTGAC	*A. thaliana DRM1*	Salk Institute (SIGnAL)	1995
drm1-RP	TTGCAGGAGCAAATATGGAAC			1996
drm1-LP1	GGAACATGGTGATAGTGATG	*A. thaliana DRM1*	This work	2138
drm1-RP1	GGTTCCATATTTGCTCCTGC			2139
drm1-LP2	GTCAGTTTGATGAGACGGCTC	*A. thaliana DRM1*	This work	2140
drm1-RP2	CATCCATCATCAATCAGACTG			2141
drm1-LP3	TCTGGATTTCATTGGTCTTACC	*A. thaliana DRM1*	This work	
drm1-RP3	GGTGTCATAGCAACGTTCTC			
drm1-LP4	AAGATCTCCAGCCACTTGTAC	*A. thaliana DRM1*	This work	
drm1-RP4	CAGACAAGTGATAAGCAACTG			
drm1-LP5	CAAATGGAATCAACGTCCTG	*A. thaliana DRM1*	This work	
drm1-RP5	ACAGAACAACTTGGGTCTTC			
drm1-LP6	TTCAGTCACATTTGCCTGCTC	*A. thaliana DRM1*	This work	
drm1-RP6	CCTAACGAAACATTGCACACAC			
drm1-LP7	TGACCGATCATATATCCTTCC	*A. thaliana DRM1*	This work	
drm1-RP7	CACAGACTGAGAAATTTGAGC			
drm1-LP8	TGCGTGTTACTAGTTGAGAGC	*A. thaliana DRM1*	This work	
drm1-RP8	TACATATGTACCCTGCGCAAG			
BP	ATTTTGCCGATTTCGGAAC	T-DNA insertion	Salk Institute (SIGnAL)	2008
drm2-LP	AGATCGCTTCCAGAGTTAGCC	*A. thaliana DRM2*	Salk Institute (SIGnAL)	1997
drm2-RP	TTGTCGCAAAAAGCAAAAGAG			1998
cmt3-LP	CCCTCAACAATTAACTGACGC	*A. thaliana CMT3*	Salk Institute (SIGnAL)	1999
cmt3-RP	ATAAGAGAAGGAGCTGCTGCC			2000
ros1-4-LP	CCAGTTAAGGACAGAACACCG	*A. thaliana ROS1*	Salk Institute (SIGnAL)	2142
ros1-4-RP	TCGTCTTTCGATCAAATCCAC			2143

3.- Plant materials and growth conditions.

Once in soil, wild-type *N. benthamiana* and wild-type and mutant *A. thaliana* Columbia ecotype (Col-0) plants were grown in a growth chamber with an 8 h photoperiod (8/16 h, light/dark) at 21ºC. After geminivirus inoculation, plants were transferred to a growth chamber with a 16 h photoperiod (16/8 h, light/dark) at 21ºC.

Tomato (*Solanum lycopersicum* cv. Moneymaker) plants were grown in a growth room with a 16 h photoperiod (16/8, light/dark, 25/20ºC), before and after geminivirus inoculation.

A. thaliana mutants were provided by the Nottingham *Arabidopsis* Stock Centre (NASC) or other groups. The following mutants have been used in this thesis: *met1-3* (AT5G49160, Dr. César Llave Correas), *ddc (drm1-2 drm2-2 cmt3-11)* (CS16384 / AT5G15380 AT5G14620 AT1G69770, NASC) and *ros1-4* (Salk_045303 / AT2G36490, Dr. Teresa Roldán Arjona). *A. thaliana* seeds were surfaced-sterilized by bleach and HCl vapors for 4 h and placed at 4ºC for 2 days in darkness for stratification. Then, seeds were sown on MS plates (4.33 g/L Murashige and Skoog basal medium, 15 g/L sucrose, 0.5 g/L morpholineethanesulfonic acid (MES); pH adjusted to 5.7) sealed with Micropore paper tape, and placed into a growth chamber with a 16 h photoperiod at 21ºC. After 9-10 days growing in vertical orientation, plants were transferred to soil. In the case of the *met1-3* seeds, the MS medium was supplemented with 15 µg/mL phosphinotricin to select for the mutation.

4.- Geminivirus infection assays.

4.1.- *N. benthamiana* infection.

Six-week-old *N. benthamiana* plants were agroinoculated with the infectious clones for (Table 1): *Tomato yellow leaf curl Sardinia virus* (TYLCSV; Accession (AC): L27708; Collection number (CN): 2419; Reina et al., 2003 (unpublished)) cloned into the binary vector pGreen-0229 (Lozano-Durán et al., 2011), and *Tomato yellow leaf curl virus* (TYLCV; AC: AJ519441; CN: 5568; Navas-Castillo et al., 2000) cloned into the binary vector pBIN19 (Navas-Castillo et al., 1999). *Agrobacterium* cultures carrying each infectious clone were grown overnight in LB broth medium plus appropriate antibiotics. Then, cultures were centrifuged (4000 rpm, 15 min) and LB medium was substituted by agroinoculation buffer (10 mM MES, 10 mM MgCl$_2$ and 100 µM acetosyringone), adjusting the optical density to 1 (OD$_{600}$=1). Following an incubation of 4 hours at room temperature in the dark, the axillary bud of the fourth/fifth leaf was inoculated with these cultures by using a syringe with needle. As control, some plants were mock-inoculated with *Agrobacterium* cultures carrying the binary vector with no viral DNA (mock). Unless otherwise stated, the second and third most apical leaves were sampled and pooled at 18 days after inoculation. Samples were processed for RNA and/or DNA isolation and subsequent analysis.

4.2.- *A. thaliana* infection.

Four to five-week-old *A. thaliana* plants were agroinoculated with the infectious clones (Table 1): *Beet curly top virus* (BCTV; AC: M24597; CN: 2543; Stanley et al., 1986) cloned into the binary vector pBINX1 (Briddon et al., 1989), *Cabbage leaf curl virus* (CaLCuV; AC: U65529 (DNA-A) and U65530 (DNA-B); CN: 6262 (DNA-A) and 6264 (DNA-B)) cloned into pGreen-0229, and the TYLCV infectious clone previously mentioned. *Agrobacterium* cultures harbouring each infectious clone were inoculated as indicated. *Agrobacterium* cultures were grown overnight in LB broth medium. By using a syringe with needle, 2-3 drops of culture were placed on the apical meristem before the flowering, at the same time that 10-20 punctures were applied around the meristem to favour the entry of the bacteria. After inoculation, plants were covered with a plastic film for 3 days. As control, plants were mock-inoculated with *Agrobacterium* cultures carrying the empty vector. Unless otherwise stated, samples were sampled at 26 days after inoculation. Apical rosette leaves and/or inflorescences were sampled. Samples were processed for DNA isolation and subsequent analysis.

4.3.- Tomato infection.

Three-week-old tomato plants (*Solanum lycopersicum* cv. Moneymaker) were infected with a clone of the isolate [ES:Alm:Pep:99] of TYLCV-Israel (TYLCV-IL, AC: AJ489258, Morilla et al., 2005) either using *Agrobacterium* or the whitefly *Bemisia tabaci*. A virus-free colony of *B.*

tabaci MED species (former Q biotype), originated from field individuals collected in Málaga (Spain), were reared on melon (*Cucumis melo* L. cv. ANC42, IHSM-UMA-CSIC seed bank) plants in cages covered by insect-proof netting. To obtain viruliferous whiteflies, non-viruliferous whiteflies were given a 24 h acquisition access period (AAP) on systemically-infected young leaves of tomato plants previously infected with TYLCV-IL by agroinoculation. Similarly, whiteflies used as non-viruliferous controls were enclosed in clip-cages attached to virus-free plants for the same 24 h period. After the AAP, whiteflies were transferred to healthy plants for a 48 h inoculation access period (IAP) (40 whiteflies per test plant in clip-on cages). For the agroinfection, the axillary bud between the third and fourth tomato leaf was infiltrated, following a procedure similar to that in *N. benthamiana* plants. As control, plants were mock-inoculated with *Agrobacterium* carrying the empty vector. TYLCV-IL infected and control plants were monitored along 21 days, and symptoms were annotated. The second most recently expanded leaf from the apex was harvested at different time points (same plants were harvested for 2 and 14 dpi and for 7 and 21 dpi). The presence of viral DNA was analysed in each test plant at 21 dpi by hybridization of tissue blots from young leaf petiole cross sections as previously described by Pereira-Carvalho, 2015), and just infected plants were further analysed. For each infection method and time point, the leaf tissue from six different infected plants was pooled and used in downstream analysis. A total of three biological replicates were processed per condition and time point. Samples were processed for RNA and DNA isolation and subsequent analysis.

5.- Virus-induced gene silencing (VIGS): agroinoculation procedure.

To carry out the virus-induced gene silencing (VIGS), different cultures of *Agrobacterium* GV3101 containing pTV00-derived constructs (in TRV2 vector) and *Agrobacterium* C58c1 carrying pBINTRA6 (TRV1 vector) (Ratcliff et al., 2001) (Table 1) were grown overnight in LB broth medium plus the appropriate antibiotics. Cultures were centrifuged (4000 rpm, 15 min) and LB medium was substituted by agroinfiltration buffer (10 mM MES, 10 mM MgCl$_2$ and 100 µM acetosyringone), adjusting the optical density to 1 (OD$_{600}$=1). Following an incubation of 4 hours at room temperature in the dark, cultures containing the TRV1 and TRV2 constructs were mixed at a 1:1 ratio. The underside of two leaves of six-week-old *N. benthamiana* plants was infiltrated with the mixed culture. In some experiments, the VIGS-treated plants were also infected with geminivirus at the same time, following the procedure described in section 4.1. Either in geminivirus-infected or uninfected VIGS-treated plants, the second and third most apical leaves were sampled and pooled at 18 days after inoculation. Samples were processed for RNA and DNA isolation and subsequent analysis.

6.- Nucleic acids molecular biology.

6.1.- Genomic DNA extraction.

In this thesis, several genomic DNA extraction protocols have been followed, depending on the objective of the experiment. All of them started macerating the plant tissue in liquid nitrogen.

6.1.1.- For geminivirus quantification by qPCR.

Unless otherwise stated, we used the so-called CTAB method (Lukowitz et al., 2000) to extract genomic DNA from geminivirus-infected plant tissue, which would be used as template for viral quantification by qPCR. A total amount of 50-100 mg of macerated tissue was homogenized into 500 µL of extraction buffer (2% cetyl trimethylammonium bromide (CTAB), 1.5 M NaCl, 100 mM Tris pH 8, 100 mM EDTA pH 8). After mixing by vortex, it was incubated for 15 min at 65ºC. Each sample was cooled at room temperature and 300 µL of chloroform/isoamilic (24/1) were added. Then, each sample was centrifuged at 13,000 rpm for 5 min and the supernatant was transferred to a new tube and mixed with 300 µL of isopropanol. Next, a new centrifugation step was carried out (13,000 rpm for 10 min) to remove the supernatant. The precipitated DNA was washed with ethanol 70% and dried for 10 min after removing the ethanol. Milli-Q H_2O was added to the pellet to dissolve it. Subsequently, 2 µL of RNase (10 mg/mL) (Invitrogen) were added to the DNA samples, which were incubated for 15 min at 37ºC.

6.1.2.- For Southern blot.

For the Southern blot, *N. benthamiana* genomic DNA was extracted by a high-quality DNA extraction method described by Dellaporta et al. (1983), using a starting tissue amount of 100 mg.

6.1.3.- For rDNA quantification by qPCR.

In order to measure by qPCR the amount of ribosomal DNA from *N. benthamiana* or *A. thaliana*, we improved the quality of the purified genomic DNA by using a method that, unlike the CTAB one, contains phenol (Phenol method). This method was also followed to purify DNA that would be used as template for Rolling Circle Amplification (RCA). A total amount of 50-100 mg of macerated tissue was homogenized into 400 µL of extraction buffer (50 mM EDTA pH 8, 100 mM Tris pH 8, 500 mM NaCl, 7 µL β-mercaptoethanol/10 mL of buffer) and 80 µL of SDS 10%. After mixing by vortex, it was incubated for 10 min at 65ºC. After cooling at room temperature, 160 µL of potassium acetate 5 M were added. The mix was homogenized by vortex, incubated for 10 min on ice and centrifuged at 13,000 rpm for 20 min. The supernatant was recovered and mixed with 0.5 volumes of phenol pH 8 and chloroform; subsequently, it was centrifuged at 13,000 rpm for 5 min and the supernatant was recovered and mixed with 1 volume of chloroform. A new centrifugation step was carried out and 700 µL of isopropanol were added to the recovered supernatant.

It was mixed and incubated for 30 min, at -20ºC, and subsequently centrifuged at 13,000 rpm for 20 min. The pellet was washed with ethanol 70% and dried for 10 min after removing the ethanol. Milli-Q H_2O was added to the pellet, which was dissolved and treated with 2 µL of RNase A (10 mg/mL) (Invitrogen). Additionally, a second procedure, named Hirt method, was used for DNA extraction and subsequent RCA, followed as described by Pradeep (1993).

6.1.4.- For BS-Seq (Bisulfite-Sequencing).

To isolate genomic DNA from tomato plants for whole-genome bisulfite sequencing (BS-Seq), we used the DNeasy Plant Mini Kit (Qiagen) with a starting amount of 100 mg, following the manufacturer's guidelines.

6.2.- Absolute quantification of VS and CS strands by two-step qPCR.

Quantification of the virion-sense (VS) and the complementary-sense (CS) strands was performed following a two-step qPCR protocol described by Rodríguez-Negrete et al. (2014). This procedure was carried out to quantify the TYLCV-IL accumulation in tomato plants. First, an extension reaction with the T4 DNA polymerase (Takara) was performed in 10 µL containing: 1 µM primer (a different primer for each strand (Table 2)), 1% BSA, 1 µM dNTPs, 1 µL T4 DNA polymerase buffer and 15 ng of DNA as template. The mix was denatured at 95ºC for 10 min and cooled at room temperature, and then, incubated at 37ºC for 30 min after addition of 1 unit of T4 DNA polymerase. Next, primers were removed using the QIAquick purification kit (Qiagen) according to the manufacturer's instructions. DNA was eluted in 100 µL of commercial water. In the second step, 2 µL of the previous reaction were supplied as qPCR template, using specific primers for VS and CS quantification (Table 2). For the quantification of total viral DNA, only the second step was carried out, using 3 ng as qPCR template and specific primers (Table 2). The absolute quantification of VS and CS strands, and total viral DNA, was obtained by extrapolation of the qPCR-derived Ct values with the corresponding standard curves. These standard curves were generated using serial dilutions containing 10^8 to 10^2 molecules of circular ssDNA or dsDNA per µL, by applying linear regression analysis of the Ct value over the $\log_{10}$ of the amount of DNA (described in Rodríguez-Negrete et al., 2014). The *25S* gene was used to normalize the TYLCV-IL DNA levels.

6.3.- Rolling Circle Amplification (RCA).

The Rolling Circle Amplification (RCA) procedure was carried out using Illustra TempliPhi 100 kit (GE Healthcare), as follows. Aliquots with 5 µL of sample buffer were mixed with 1 µL of each DNA sample (1-50 ng). Next, the sample was denatured by heat at 95ºC for 3 min and then cooled at 4ºC. A master mix containing 0.2 µL of enzyme mix (it includes Phi29 DNA polymerase) and 5 µL of reaction buffer per each reaction was added to the cooled denatured

sample. It was incubated overnight at 30ºC and then, heat-inactivated at 65ºC for 10 min. As a positive control of amplification, 1 µL of pUC vector (included in the kit) was used as template.

6.4.- RNA extraction.

In this thesis, several RNA extraction methods have been followed, depending on the objective of the experiment. All of them started macerating the plant tissue in liquid nitrogen.

6.4.1.- For RT-sqPCR and RT-qPCR.

Unless otherwise stated, total RNA was extracted as follows. A total amount of 50-100 mg of macerated tissue was homogenized into 400 µL of extraction buffer (100 mM LiCl, 10 mM Tris-HCl pH 8, 10 mM EDTA, 1% SDS) plus 400 µL of phenol pH 8 and 400 µL of chloroform. The mix was centrifuged at 13,000 rpm for 15 min. The supernatant was recovered and mixed with 1/3 volumes of LiCl 8M. It was incubated overnight at 4ºC. After incubation, it was centrifuged at 13,000 rpm for 15 min, at 4ºC. The pellet was resuspended in 200 µL of RNase-free H_2O, and 1/10 volume of sodium acetate 3M pH 5.2 and 2.5 volumes of ethanol 100% were added. The mix was incubated for 20 min at -20ºC and centrifuged at 13,000 rpm for 15 min, at 4ºC. The pellet was resuspended in ethanol 70% and a new centrifugation step was performed to remove the ethanol. Then, it was dried and 30 µL of RNase-free H_2O were added. A total RNA amount of 5-8 µg was treated with RNase-free Turbo DNaseI (Ambion) according to the manufacturer's guidelines, and cleaned up by basic phenol/chloroform treatment. Total and cleaned up RNA concentration and quality was determined by using the spectrophotometer NanoDrop 8000 (ThermoFisher Scientific).

6.4.2.- For RNA-Seq and sRNA-Seq.

Total RNA extracted from tomato plants was used for mRNA and small RNA (sRNA) sequencing, so that we changed the extraction method due to the previous one eliminates the sRNA fraction. We followed an adapted protocol using Trizol reagent (Invitrogen) to yield total RNA amounts higher than 30 µg. Using a 15 mL falcon, we homogenized 1 g of tissue and 10 mL of Trizol and incubated for 5 min at room temperature. Subsequently, 2 mL of chloroform were added and the mix was incubated for 3 min else. It was centrifuged at 9,000 rpm for 10 min, at 4ºC, and the aqueous phase was recovered in a new falcon; the chloroform step was repeated until the interphase disappeared. One volume of cold isopropanol was added and mixed; then, it was incubated for 20 min at room temperature and centrifuged at 9,000 rpm for 15 min, at 4ºC. The pellet was washed with ethanol 75% and new centrifugation step was carried out to eliminate the ethanol. The pellet was dried for 10 min and dissolved in 200 µL of RNase-free H_2O. A total RNA amount of 40 µg was treated with RNase-free Turbo DNaseI (Ambion) according to the manufacturer's

guidelines, and cleaned up by basic phenol:chloroform treatment. Total and cleaned up RNA concentration and quality was determined using the Agilent 2100 Bioanalyzer.

6.5.- cDNA synthesis.

For cDNA synthesis, 300-500 ng of cleaned up RNA were used as template for the reaction. We used the iScript cDNA Synthesis Kit (Biorad), following the instructions provided by the manufacturer.

6.6.- Retrotranscription-semi-quantitative PCR (RT-sqPCR).

To measure the accumulation of precursor ribosomal RNA (pre-rRNA), cleaned up RNA from *N. benthamiana* apical leaves was converted to cDNA and amplified by standard PCR in one step, by using the SuperScript III One-Step RT-PCR system with Platinum *Taq* DNA polymerase (Invitrogen) according to the manufacturer's instructions. Per each sample, different amounts (400, 100 and 10 ng) of RNA were used as template for the RT-sqPCR procedure. Specific primers complementary to the pre-rRNA and, as a load control, the *Rubisco* mRNA (Table 2) were used for the cDNA synthesis and subsequent PCR. RT-PCR conditions for the pre-rRNA were 45ºC for 30 min; then, 30 cycles of 95ºC for 15 s, 55ºC for 30 s, 72ºC for 30 s. RT-PCR conditions for the *Rubisco* mRNA were 45ºC for 30 min; then, 30 cycles of 95ºC for 15 s, 58ºC for 20 s and 72ºC for 15 s.

6.7.- Quantitative PCR (qPCR).

For quantitative PCR (qPCR), an amount of 1-10 ng of genomic DNA or cDNA was used as template. The reaction mixture contained 2 µL of template, a primer mix (10 µM each one) and 5 µL of SsoFast EvaGreen Supermix (Biorad) in a total volume of 10 µL. The PCR conditions were: 10 min at 95ºC and 40 cycles of 15 s at 95ºC and 10 s at 60ºC. The reaction was carried out by using a CFX96 Touch Real-Time PCR Detection System instrument (Biorad). Per each sample, two or three technical replicates were performed. The final value obtained for each sample was estimated by applying the Livak's data analysis method to the Ct value obtained with primers complementary to the target gene (Table 2), and the Ct value obtained with primers complementary to the normalizer gene (primers for *AtActin, NbActin, NbEF1α,* and *Nb25S* (Table 2)).

In order to confirm the absence of contaminant DNA prior to the cDNA synthesis, a qPCR reaction using extracted RNA as template and primers for *AtActin, NbEF1α* and *Nb25S* (Table 2) was carried out.

6.8.- Southern blot: probe generation and DNA hybridization.

The probe used for the genomic DNA hybridization was generated and labelled as follows. A PCR was performed using *N. benthamiana* genomic DNA as template, and specific primers that amplify a region (801 bp) of the *Nb25S* ribosomal gene (25S2-F and 25S1-R, Table 2). The PCR conditions were: 94ºC for 5 min, followed by 32 cycles at 94ºC for 30 s, 55ºC for 30 s and 72ºC for 90 s, and followed by 7 min at 72ºC. The PCR product was A/T cloned into the pGEM-T easy vector (Promega) to yield pGENb25S, and confirmed by sequencing with the SP6 and T7 commercial primers. The pGENb25S construct was used as template for a new PCR with the 25S2-F and 25S1-R primers and the PCR DIG Labelling Mix (Roche) (it contains digoxigenin-dNTPs), following the manufacturer's instructions. The amplified and digoxigenin-labelled fragment was used for the hybridization.

An amount of 5-8 μg of genomic DNA from *N. benthamiana* apical leaves was run on a 0.8% agarose gel at 70 V for 4-5 h, and stained with EtBr for visualization. Additionally, 5 μg of genomic DNA from *N. benthamiana* apical leaves were digested with 6 μL of *Eco*RI in a final volume of 120 μL, for 5 h at 37ºC. The DNA was precipitated with NaOAc 3M, 1 μL glycogen and 2 volumes of 100% ethanol. It was incubated at -20ºC for 20 min and centrifuged for 20 min at 8,000 rpm. The pellet was washed with 500 μL of 70% ethanol, and resuspended in milli-Q H_2O. The DNA was run on a 0.8% agarose gel at 70 V for 4-5 h. The agarose gel was treated with HCl 0.25 M for 15 min and washed two times with distilled water. Then, it was treated with denaturing buffer (1.5 M NaCl, 0.5 M NaOH) for 30 min, and subsequently treated with neutralizing buffer (3 M NaCl, 0.5 M Tris pH 7.2, 0.5 M EDTA pH 8) for 30 min. The DNA transfer was done by setting a Whatman paper (previously sunken into SSC 10x buffer (3 M NaCl, 0.3 M $C_6H_5Na_3O_7$)), the agarose gel, a Nylon membrane (Sigma, 0.45 μm pore size), three Whatman papers and 20-30 paper towels under a weight, taking place overnight. The membrane was cross-linked using UV light at 1,200 J/m^2. Then, it was prehybridized with the corresponding buffer (SSC 5x, 1% blocking reagent, 0.1% N-lauryl sarcosine (w/v), 0.1% NaCl (w/v), 0.02% SDS (w/v)) for 1 hour at 65ºC. The hybridization with the probe (denatured and diluted into prehybridization buffer) was carried out at 65ºC, overnight. Next, the membrane was rinsed with SSC 2x (0.3 M $NaC_6H_5O_7$ and 0.03 M NaCl) and 0.1% SDS, and incubated for 5 min at room temperature. Then, it was incubated three times with SSC 0.5x and 0.1% SDS for 10 min at 65ºC, and washed with Washing buffer (70 mM maleic acid, 150 mM NaCl and 0.3% Tween 20) for 5 min. The membrane was incubated with Buffer 2 (70 mM maleic acid, 150 mM NaCl and 1% blocking reagent (w/v)) for 30 min, and then, with Buffer 2 containing the anti-digoxigenin antibody solution (dilution 1/10,000) for 1 hour, at room temperature. The membrane was washed twice with Washing buffer for 15 min, and Buffer 3 (0.1 M Tris, 10 mM NaCl, 50 mM $MgCl_2$) for 5 min, at room temperature. Finally, the membrane was incubated with Buffer 3 containing CSPD (dilution 1/10,000) for 5 min at room temperature. The CSPD is a chemiluminescent substrate for alkaline phosphatase that enables DNA detection by producing light emission, which can be recorded on luminescence imager systems.

7.- Libraries construction and sequencing.

Performed by CNAG (Centro Nacional de Análisis Genómico, Barcelona, Spain).

RNA-Seq libraries were generated and sequenced at CNAG (Centro Nacional de Análisis Genómico, Barcelona, Spain). Total RNA was assayed for quantity and quality using Qubit® RNA HS Assay (Life Technologies) and RNA 6000 Nano Assay on a Bioanalyzer 2100. The RNASeq libraries were prepared from total RNA using the TruSeq®Stranded mRNA LT Sample Prep Kit (Illumina Inc., Rev.E, October 2013). A total of 24 libraries were constructed (12 for each inoculation method; 3 biological replicates at 2, 7, 14 and 21 dpi). The libraries were sequenced on HiSeq2000 (Illumina, Inc) in paired-end mode with a read length of 76bp using the TruSeq SBS Kit v4. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 1.18.64 or 1.18.66.3) and followed by generation of FASTQ sequence files by CASAVA.

sRNA libraries were generated and sequenced at CRG (Centre For Genomic Regulation, Barcelona, Spain). Libraries were prepared using the Illumina TruSeq small RNA sample prep kit according to the manufacturer's instructions and were validated on an Agilent 2100 Bioanalyzer using a DNA High Sensitivity chip and quantified by qPCR using the Kapa Library Quantification kit for Illumina (Roche). Sequencing was performed on an Illumina HiSeq 2500 using 50 bp single reads with HiSeq v4 sequencing chemistry. A total of 12 high-coverage libraries were generated and sequenced (6 for each inoculation method; 2 biological replicates at 7, 14 and 21 dpi).

Bisulfite sequencing (BS-Seq) library construction was performed at CNAG. Total genomic DNA (2 μg) was spiked with unmethylated bacteriophage λ DNA (5 ng of λ DNA per microgram of genomic DNA). The DNA was sheared to 50–500 bp in size using Covaris LE220 ultrasonicator, and fragments of 150–300 bp were size-selected using AMPure XP beads (Agencourt Bioscience). The libraries were constructed using the KAPA Library Preparation Kit with no PCR Library Amplification/Illumina series (Roche-Kapa Biosystems) together with the Illumina single index adaptors (Illumina). After adaptor ligation, the DNA was treated with sodium bisulfite using the EpiTect Bisulfite kit (Qiagen) following the manufacturer's instructions. Enrichment for adaptor-ligated DNA was carried out through seven PCR cycles using KAPA HiFi HotStart Uracil+ReadyMix PCR 2x Kit (Roche-Kapa Biosystems). Four libraries were constructed (2 for each inoculation method; 2 biological replicates at 14dpi). Library quality was monitored using the Agilent 2100 Bioanalyzer DNA 7500 assay, and the library concentration was estimated using quantitative PCR using the KAPA Library Quantification Kit for Illumina® Platforms, v1.14 (Roche-Kapa Biosystems). Paired-end DNA sequencing (2×101bp) of the converted libraries was performed using the HiSeq2000 (Illumina) following the manufacturer's protocol with HiSeq Control Software (HCS) 1.5.15.1 to achieve no less than 24x tomato genome coverage. Images analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 1.13.48) and followed by generation of FASTQ sequence files.

8.- Bioinformatic analysis.

Performed by CNAG (Centro Nacional de Análisis Genómico, Barcelona, Spain), Dr. Chen Jiao and Dr. Zhangjun Fei (Cornell University, USA), and Álvaro Piedra Aguilera.

RNA-Seq paired-end reads were mapped against the TYLCV reference genome (Genbank Acc. No. AJ489258) using STAR version 2.5.1b (Dobin et al., 2013) with ENCODE parameters for long RNA. Viral transcriptional units were quantified using RSEM version 1.2.28 (Li and Dewey, 2011) with default parameters. To make sense/antisense coverage plots, we extracted alignments from both strands using coverageBed (Quinlan and Hall, 2010) and plotting it using the GViz R package.

For sRNA data adaptor and ambiguous bases were first trimmed from the raw sRNA reads using a custom Perl script. Reads longer than 15 bp were then aligned to a combined RNA database consisting of tRNA (transfer RNA), snoRNA (small nucleolar RNA), snRNA (small nuclear RNA) and rRNA (ribosomal RNA), sequences of plants, and those could be aligned were discarded. The resulting cleaned reads were mapped to TYLCV genome, respectively, using Bowtie (Langmead et al., 2009), allowing one mismatch. The sRNA size distribution was summarized for both origins. The total number of reads, reads in forward and reverse orientation and reads starting with A, C, G and T for TYLCV sRNA size classes from 20 to 25 nt, were estimated using MISIS (Seguin et al., 2016). Plotting of vsRNAs along the viral genome was accomplished by Excel or MISIS.

From the BS-Seq data, potential PCR duplicates were removed: read pairs having identical bases at positions from 10 to 80 in both left and right reads were defined as duplicated pairs and then collapsed into unique read pairs. The resulting reads were further processed to remove adaptor and low-quality sequences using Trimmomatic (Bolger et al., 2014). The trimmed reads were then mapped to TYLCV genome genomes using a methylation-aware aligner Bismark v0.17.0 (--bowtie1 -n1) (Krueger and Andrews, 2011). The methylation information was then extracted from the alignments by a script "bismark_methylation_extractor" provided in Bismark and the resulted cytosine reports were separated according to the cytosine context, CG, CHG and CHH. Plotting of the viral methylome was accomplished using ggplot2 or Excel.

Aims

The main aims of this thesis are:

1. To gain insight into the connection between virus-derived small RNAs (vsRNAs), viral genome methylation and gene expression in the pathosystem *Tomato yellow leaf curl virus* (TYLCV) – tomato.

2. To determine the biological relevance of the plant DNA methylation machinery during geminivirus infection in *Arabidopsis thaliana* and *Nicotiana benthamiana.*

3. To assess if geminiviruses are transmitted to the offspring via seeds in *Arabidopsis thaliana*.

Chapter 1

Integrated single-base resolution maps of transcriptome, sRNAome and methylome of *Tomato yellow leaf curl virus* (TYLCV) in tomato

The content of this Chapter has been entirely published in Scientific Reports (*Scientific Reports,* February 27th, 2019. **Vol.9**, Article number: 2863). The original manuscript is included as appendix I.

BACKGROUND

Geminiviruses must confront post-transcriptional and transcriptional gene silencing (Raja et al., 2010; Pooggin, 2013; Hanley-Bowdoin et al., 2013, Ramesh et al., 2017). Several works have deep-sequenced vsRNA populations from different geminivirus-infected plants such as tomato (Donaire et al., 2009; Yang et al., 2011a; Miozzi et al., 2013; Xu et al., 2017), *Nicotiana benthamiana* (Yang et al., 2011a), *Arabidopsis thaliana* (Aregger et al., 2012) or cassava (Rogans et al., 2016). Although the results slightly vary depending on the geminivirus-host assayed, generally 21, 22 and 24-nt vsRNAs are the most abundant size classes. Geminiviral sRNAs are produced from both viral strands, cover almost the entire circular viral genome and accumulate at elevated levels in several hotspot regions. There are two reports on TYLCV vsRNAs generated during the infection in tomato that provide limited information, since the data came from a reduced number of reads (1212 reads) (Donaire et al., 2009), or belonged to field-collected tomato samples infected in most cases, with other tomato viruses (Xu et al., 2017).

Geminivirus double-stranded DNA (dsDNA) associates with host histones to form minichromosomes (Pilartz and Jeske, 1992; 2003; Pooggin, 2013), which carry the same epigenetic marks as the host chromatin, i.e. DNA methylation at cytosines and histone modifications (Raja et al., 2008; Castillo-González et al., 2015; Ceniceros-Ojeda et al., 2016). The level of geminiviral DNA methylation in infected plants depends on the geminivirus-host analysed, ranging from 50% to 2% (Bian *et al.*, 2006; Raja *et al.*, 2008; Rodríguez-Negrete *et al.*, 2009; Zhang *et al.*, 2011; Yadav & Chattopadhyay, 2011; Yang *et al.*, 2011b; Paprotka *et al.*, 2011; Sahu *et al.*, 2014; Ceniceros-Ojeda *et al.*, 2016; Torchetti *et al.*, 2016; Jackel *et al.*, 2016; Deuschle *et al.*, 2016). In general, low levels of DNA methylation are reported for geminivirus, which has led to the hypothesis that geminiviruses evade repressive cytosine methylation and TGS very efficiently by their resourceful replication mechanisms (Paprotka et al., 2011; Pooggin, 2013). On the other hand, in tissues showing host recovery, an increase of viral DNA methylation levels or minichromosome compaction is detected, supporting an active role of TGS in host defence (Raja et al., 2008; Rodríguez-Negrete et al., 2009; Raja et al., 2014; Ghoshal and Sanfaçon, 2015, Ceniceros-Ojeda et al., 2016; Jackel et al., 2016).

To obtain insights into the connection between virus-derived small RNAs (vsRNAs), viral genome methylation and gene expression we have generated the transcriptome, sRNAome and methylome from the geminivirus TYLCV-infected tomato plants. In the current study, using Illumina high-throughput sequencing technology, we have obtained the transcriptome (mRNA-Seq) and sRNAome (sRNA-Seq) of TYLCV at different time points during the infection, and compared two viral inoculation methods: the "artificial" but widely-used agroinoculation and the infection by its natural vector, the whitefly *B. tabaci*. Moreover, using bisulfite treatment and deep sequencing (BS-Seq), the viral methylome was analysed at a time point in which the infection was well stablished, providing a reliable landscape for TYLCV methylation from tomato plants. Using this approach, we have generated single-nucleotide resolution integrated transcriptomic and epigenetic maps for TYLCV. This work represents the deepest molecular characterization at a single-nucleotide resolution, so far depicted for a geminivirus.

RESULTS

1.1. TYLCV accumulation and gene expression in tomato is similar regardless of the inoculation method.

Three-week-old tomato plants were infected with TYLCV by agroinoculation or the whitefly *Bemisia tabaci (*referred from now on as *Bemisia*). Plants exposed to aviruliferous whiteflies or *Agrobacterium* carrying an empty binary plasmid were used as negative controls. TYLCV symptoms appeared at 7 and 10 days post-inoculation (dpi) in agroinfected and whitefly-infected plants, respectively. At 14 dpi, all inoculated plants showed TYLCV symptoms, although these were slightly more advanced in agroinfected plants, and at 21 dpi all plants displayed severe symptoms. Apical leaf tissue was sampled at 2, 7, 14 and 21 dpi and three independent samples at each time point were collected to generate three biological replicates. DNA was extracted to quantify the accumulation of total viral DNA and virion-sense (VS) and complementary-sense (CS) strands (Rodríguez-Negrete et al., 2014).
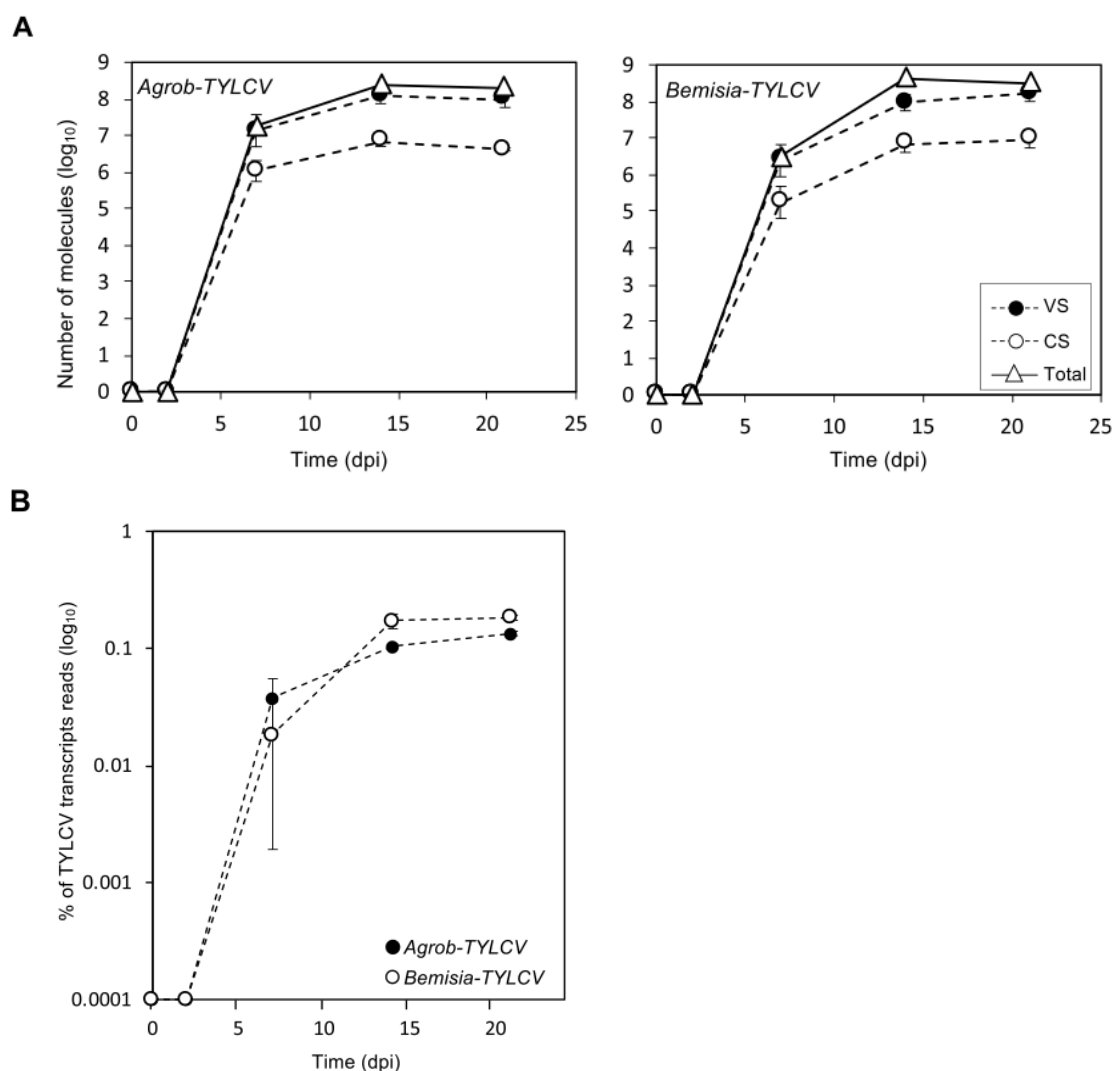

Figure 1. Genome and transcripts reads accumulation from TYLCV. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or using *B. tabaci* (*Bemisia-TYLCV*). Data from different time points during the

systemic infection are shown (2, 7, 14 and 21 dpi). **(A)** Absolute quantification of virion-sense (VS) or complementary-sense (CS) strands and total viral DNA of TYLCV. Total DNA was analysed by two-step qPCR to quantify VS or CS strands. Total viral DNA was quantified by standard qPCR. Values represent the average ± SE of viral molecules per ng of DNA from the three biological replicates ($\log_{10}$). **(B)** Percentage of transcripts reads that mapped to TYLCV genome relative to total RNA-Seq transcripts reads. Data are represented by the mean ± SE. Values are the mean of the three biological replicates.

No viral DNA was detected in the mock plants from both inoculation methods. The pattern of viral DNA accumulation was similar in infected plants treated with *Agrobacterium* or *Bemisia*, with the total amount of viral DNA increasing exponentially before reaching a *plateau* at 14 dpi at a similar level; $2.8 \times 10^8 \pm 1.1 \times 10^8$ and $4.3 \times 10^8 \pm 0.9 \times 10^8$ molecules per ng of extracted DNA, respectively (Figure 1A). In accordance with the symptoms development, at 7dpi, whitefly-treated plants showed fewer viral molecules ($5.4 \times 10^6 \pm 3.9 \times 10^6$ per ng) than plants infected by agroinfiltration ($2.1 \times 10^7 \pm 0.8 \times 10^7$ per ng) (Figure 1A). Consistent with previous reports for other geminiviruses (Rodríguez-Negrete et al., 2014), VS strands accumulated to higher levels than CS strands (from 13 to 22 times) at all time-points during the infection (Figure 1A, Table S1).

To further characterize the infection, the viral mRNA landscape was determined. Total RNA was extracted from the same samples used for viral quantification and subjected to RNA-Seq. On average, 30 million 75 bp paired-end reads per sample (range, 25-56 million) were obtained (Table S2). About 0.13% to 0.18% of the reads per sample were mapped to the TYLCV genome at 21 dpi (Figure 1B, Table S3). The kinetics of TYLCV mRNA accumulation was in accordance with the amount of viral DNA previously detected. As expected, at 2dpi almost no viral transcripts were found (Figure 1B, Table S3) and a clear increase was observed in samples collected at 7 dpi from both *Agrobacterium* and whitefly-treated plants, representing 0.04±0.017% and 0.02±0.016% of the total reads, respectively. At 14 dpi, the number of viral transcripts raised approximately 3 times for *Agrobacterium*- and 9 times for *Bemisia*-treated plants, reaching a maximum level of 0.10±0.005% and 0.17±0.025% of the total reads, respectively. Those percentages were maintained at 21 dpi (Figure 1B, Table S3). No differences in viral mRNA accumulation dynamics was observed when comparing samples from whitefly- and *Agrobacterium*-mediated TYLCV infection.

To determine the distribution of the viral transcripts accumulated during the infection, RNA-Seq reads were mapped against the viral genome. The TYLCV genome (2,781 nt presented in a linear form, setting the first nucleotide of the IR to 1 in Figure 2) includes six ORFs distributed in two transcriptional units (R and L) separated by the intergenic region (IR) and transcribed from the VS or the CS strand, respectively. Our data indicated that viral mRNA reads were not homogeneously distributed along the genome and mapped according to the predicted ORFs with almost no reads mapping at the IR or at the antisense strands of the L and R transcriptional units (Figure 2, Table 1). Transcripts from the VS strand accumulated to a higher level throughout the infection than transcripts from the CS strand in both infection methods. The reads in the VS strand (V1 and V2) represented around 75% of the total viral RNA-Seq reads at any time during the infection. Although transcription of both ORFs has been proposed to occur from the same

promoter as a single transcriptional unit, the reads that mapped in the 5′ORF of V2 were less abundant than those corresponding to the V1 ORF (Figure 2). The scenario was more complex for the four ORFs located at the CS strand (C1, C2, C3 and C4). The amount of reads for the C2 and C3 ORFs was significantly higher than the amount for the C1/C4 ORFs (Figure 2). No significant differences in transcript accumulation levels of the different viral ORFs were detected among the samples collected at different times during the infection from plants treated with either *Bemisia* or *Agrobacterium.*
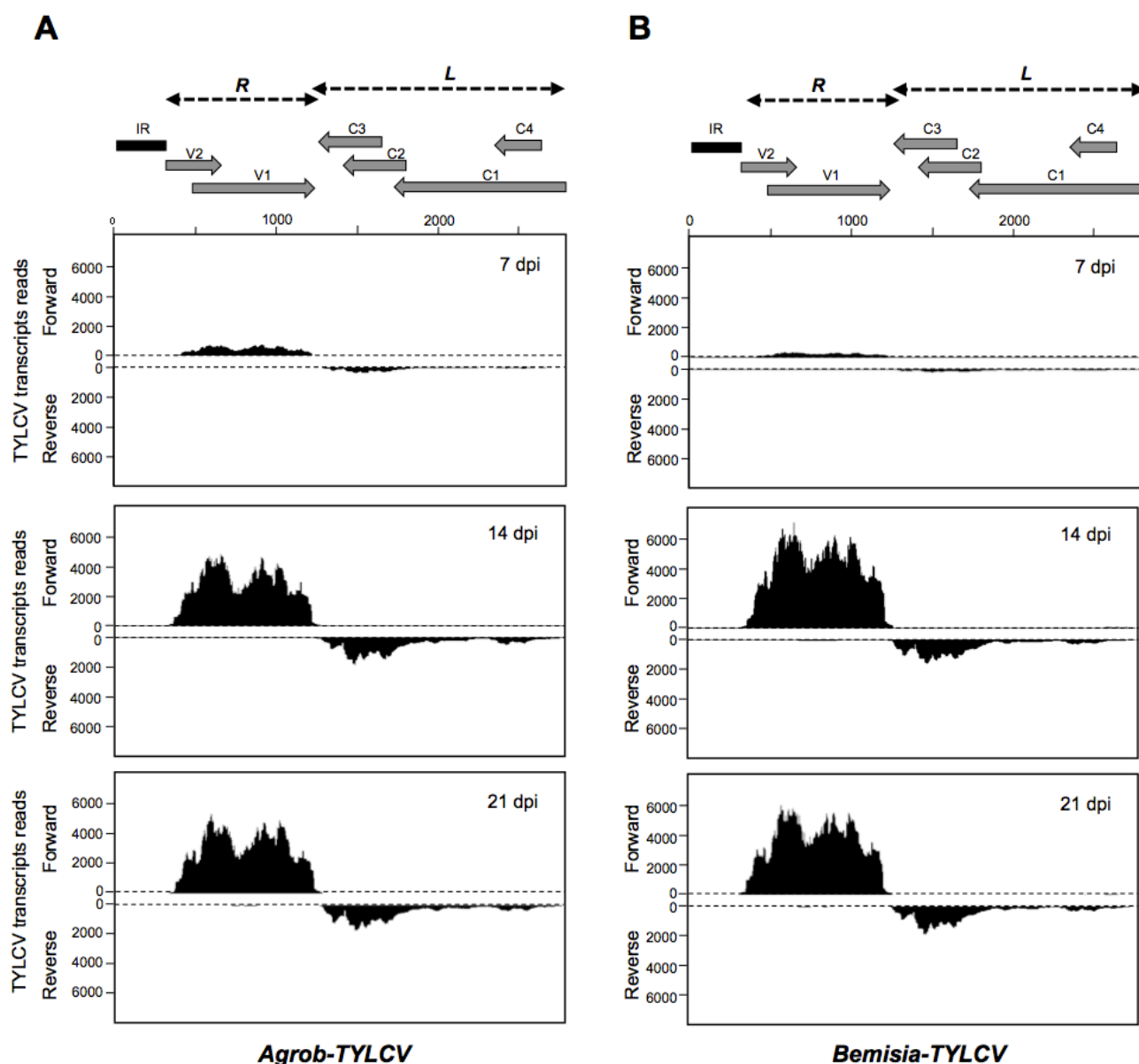
Figure 2. Maps of total transcript reads from TYLCV. Tomato plants were infected by either **(A)** agroinoculation (*Agrob-TYLCV*) or using **(B)** *B. tabaci* (*Bemisia-TYLCV*). The genome organization of TYLCV is shown schematically at the top. The predicted ORFs are shown as grey arrows and the IR as a dark rectangle. The R and L transcriptional units are indicated as black doted arrows. Data from different time points during the systemic infection are shown (7, 14 and 21 dpi). The graphs plot the number of viral transcripts reads on TYLCV genome (2781 nt). Bars above or below the axis represent sense (forward) or antisense (reverse) reads, respectively. One representative biological replicate is shown (replicate number is indicated in Table S2).

TYLCV transcripts reads per strand (%)

	Agrob-TYLCV			Bemisia-TYLCV		
	7 dpi	**14 dpi**	**21dpi**	**7 dpi**	**14 dpi**	**21dpi**
VS strand (R)	72.2±21.4	74.7±3.4	75.8±4.7	72.5±39.8	82.3±7.9	80.0±2.8
CS strand (L)	27.8±6.0	25.3±1.3	24.3±0.7	27.5±12.7	17.7±1.1	20.0±0.9

Table 1. Percentage of TYLCV transcripts reads that mapped to the virion-sense (VS) or complementary-sense (CS) strands. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or by *B. tabaci* (*Bemisia-TYLCV*). The average percentage (%) ± SE from the three biological replicates is shown at different time points during the systemic infection (7, 14 and 21 dpi). R and L viral transcriptional units are indicated.

1.2. 21, 22, and 24-nt TYLCV vsRNAs preferentially accumulate in tomato plants.

Deep sequencing of sRNAs was performed on two of the three biological replicates from the TYLCV-infected plants used to determine the transcriptome at 7, 14 and 21 dpi (Table S2). Samples at 2 dpi were not analysed due to the low amount of virus at that time point (Figure 1A, Table S1). The total number of raw sRNA reads ranged from 66 to 82 million (74 million on average, Table S2). As expected, the normalized amount of total (redundant) 20-25 nt sequences that mapped to TYLCV genome (vsRNA) increased along the infection, representing more than 5% of the total sRNA reads at 21 dpi (Figure 3A). The kinetic of vsRNA accumulation was very similar to that observed for the viral DNA and mRNA. Although vsRNA were detected at 7dpi in *Bemisia-* and *Agrobacterium*-treated plants, the levels of the latter were remarkably higher. The amount of vsRNAs strongly increased from 7 to 14 dpi (approximately 10 and 30 time-fold in *Agrobacterium* and whitefly-mediated TYLCV infection, respectively) and reached the highest accumulation at 21 dpi. No significant differences were detected between the two biological replicates of each infection method (Figure 3A).

The level of redundancy of vsRNAs increased along the infection. At 7dpi, low levels of redundancy were observed and the vast majority of the vsRNAs were read 1 to 10 times (around 77% in agroinfected plants and 89% in *Bemisia* TYLCV-infected plants, Table 2). As infection progressed, vsRNA that were read from 10 to 1000 times represented more than 42% of the total vsRNA, showing similar percentages at 14 and 21 dpi for plants infected by both methods (Table 2). The vsRNA size distribution of redundant vsRNA was similar to that previously reported in tomato plants infected with other monopartite begomovirus, such as *Tomato yellow leaf curl Sardinia virus* (TYLCSV), *Tomato yellow leaf curl China virus* (TYLCCNV) or TYLCV (Yang et al., 2011a; Miozzi et al., 2013; Xu et al., 2017). Our data showed that vsRNAs of 21 and 22-nt were the most abundant classes, representing approximately 3.5% of the total sRNA reads at 14 and 21 dpi (Figure 3B, Table S4) and around 80% from the pool of 20-25 nt vsRNA reads at any time point studied (Table S5). VsRNAs of 20, 23 and 24-nt represented more than 15% of the 20-25 nt vsRNA population, being the 24-nt the most abundant among them (around 8%, Table S5). The size distribution pattern of the redundant vsRNAs was maintained at all time points during the infection and no significant differences were detected between plants treated with the *Bemisia*

or *Agrobacterium*.
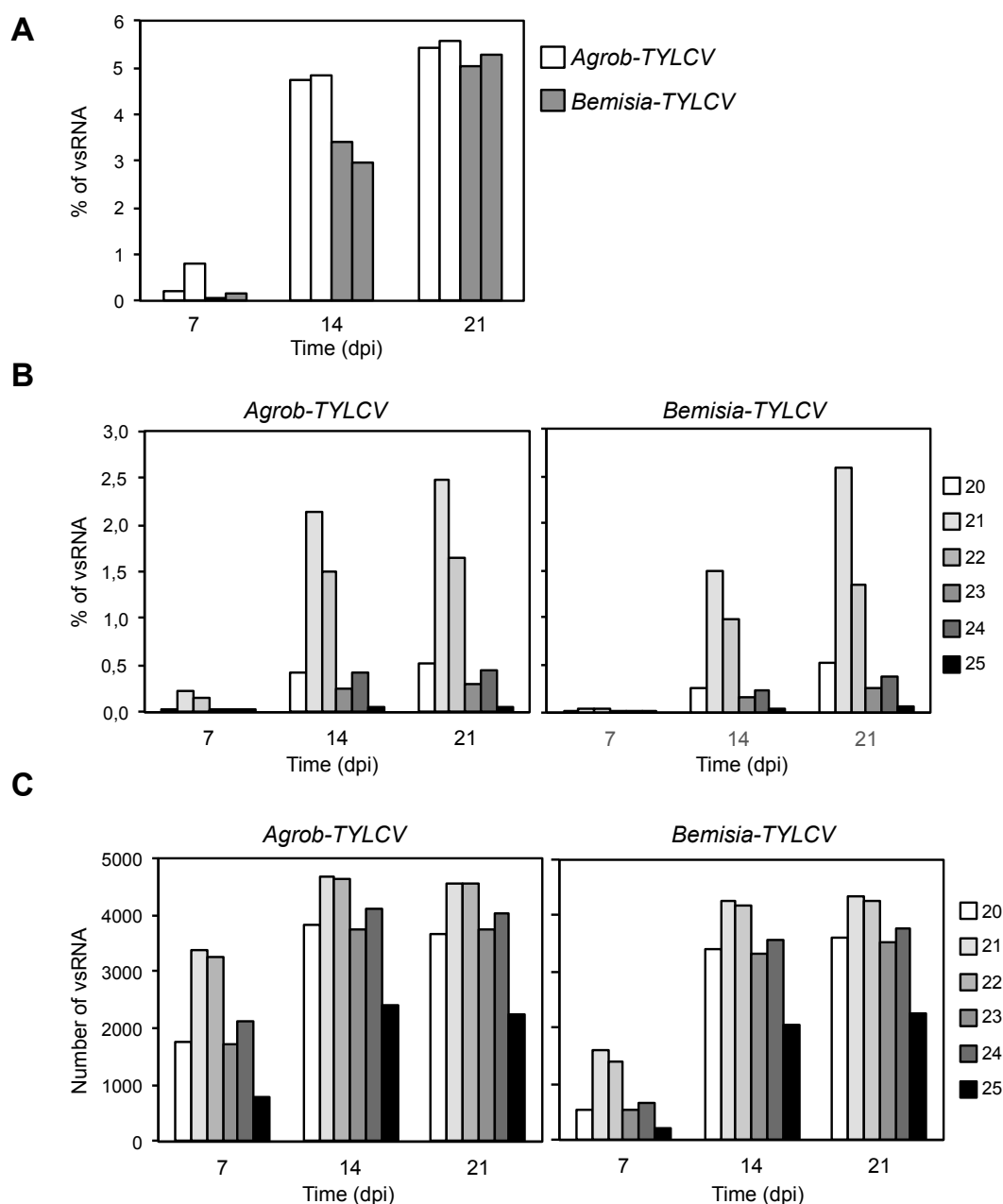
Figure 3. Accumulation and size distribution from TYLCV vsRNAs. Tomato plants were infected by either agroinoculation (*Agrob-TY*LCV) or by *B. tabaci* (*Bemisia-TYLCV*). Data from different time points during the systemic infection are shown (7, 14 and 21 dpi). **(A)** Percentage of vsRNA in the pool of 20-25 nt reads mapped to TYLCV genome relative to the total sRNA reads (19-35 nt). Each bar corresponds to an independent biological replicate. **(B)** Percentage of each size-class of 20-25 nt TYLCV vsRNA reads (redundant) relative to the total sRNA reads (19-35 nt). The bar from each vsRNA size corresponds to the mean of the two independent biological replicates. **(C)** Total number of each size-class in the pool of 20-25 nt unique TYLCV vsRNA reads. The bar from each vsRNA size corresponds to the mean of the two independent biological replicates.

The scenario was different for the unique vsRNA population (Figure 3C). Although 21 and 22-nt vsRNAs were also the most abundant classes at all times, their relative proportion was reduced as the infection progressed. At 7 dpi, they represented 50 to 60% of all unique vsRNAs, but this value decreased down to 40% in later times when the infection was well established

(Table S5). The percentages for unique 20, 23 and 24-nt vsRNAs were very similar between them and throughout the infection (Table S5). As 24-nt vsRNA is the third most represented size class of the redundant viral derived sRNAs at any time point (ranging from 6.9 to 8.8%), we can conclude that the redundancy of 24-nt vsRNAs was higher compared to the 20-nt or the 23-nt classes (Table S5). Altogether our data indicated that most of the diversity of vsRNA already exists at 7dpi. The significant increase of vsRNAs observed between 7 and 14 dpi was largely due to the production of additional copies of the same vsRNA, rather than the generation of new unique vsRNA. This redundancy increase was mainly caused by the rise in the total number of 21 and 22-nt size classes, although all the vsRNA sizes increased in a similar proportion during the infection (Table S5, Table S6).

	% vsRNA (20-25 nt)					
	Agrob-TYLCV			*Bemisia-TYLCV*		
Number of counts per vsRNA	**7 dpi**	**14 dpi**	**21 dpi**	**7 dpi**	**14 dpi**	**21 dpi**
1 - 10 $_{VS}$	37.35	27.31	26.97	42.04	28.35	26.47
1 - 10 $_{CS}$	40.04	27.41	27.58	47.21	29.05	26.64
10 - 100 $_{VS}$	9.25	16.12	15.93	4.63	15.76	16.42
10 - 100 $_{CS}$	10.10	17.13	16.77	5.26	16.33	17.42
100 - 1000 $_{VS}$	1.43	4.92	5.29	0.43	4.35	5.36
100 - 1000 $_{CS}$	1.66	5.57	5.74	0.41	4.96	5.81
>1000 $_{VS}$	0.09	0.70	0.79	0.01	0.57	0.89
>1000 $_{CS}$	0.08	0.85	0.93	0.01	0.63	1.00
Total-Unique vsRNA	13050	23424	22768	7366	20813	21813

Table 2. Redundancy of TYLCV vsRNAs. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or using *B. tabaci* (*Bemisia-TYLCV*). Data from different time points during the systemic infection are shown (7, 14 and 21 dpi). Unique 20-25 nt vsRNAs were classified according to the strand where they mapped (VS or CS) and to the number of counts, in intervals 1 to 10, 10 to 100, 100 to 1000 and >1000 reads. The values indicate the percentage (%) of vsRNA according to the number of counts on each interval using the average from the two biological replicates. The total number of unique vsRNAs (20-25 nt) at each dpi is shown.

1.3. vsRNAs distribute along both strands of the entire viral genome and accumulate in several large hotspot regions.

To examine the genomic distribution of the virus-derived sRNAs, the reads from TYLCV-infected plants were plotted against the TYLCV genome. Analysis of single-nucleotide resolution maps of 20-25 nt vsRNAs revealed that they covered both CS and VS strands of the entire circular viral genome in an approximately equal ratio at 7, 14 and 21 dpi and regardless of the infection method used (Figure 4A, Table S6), indicating that vsRNAs were generated from dsRNA precursors derived from both strands.
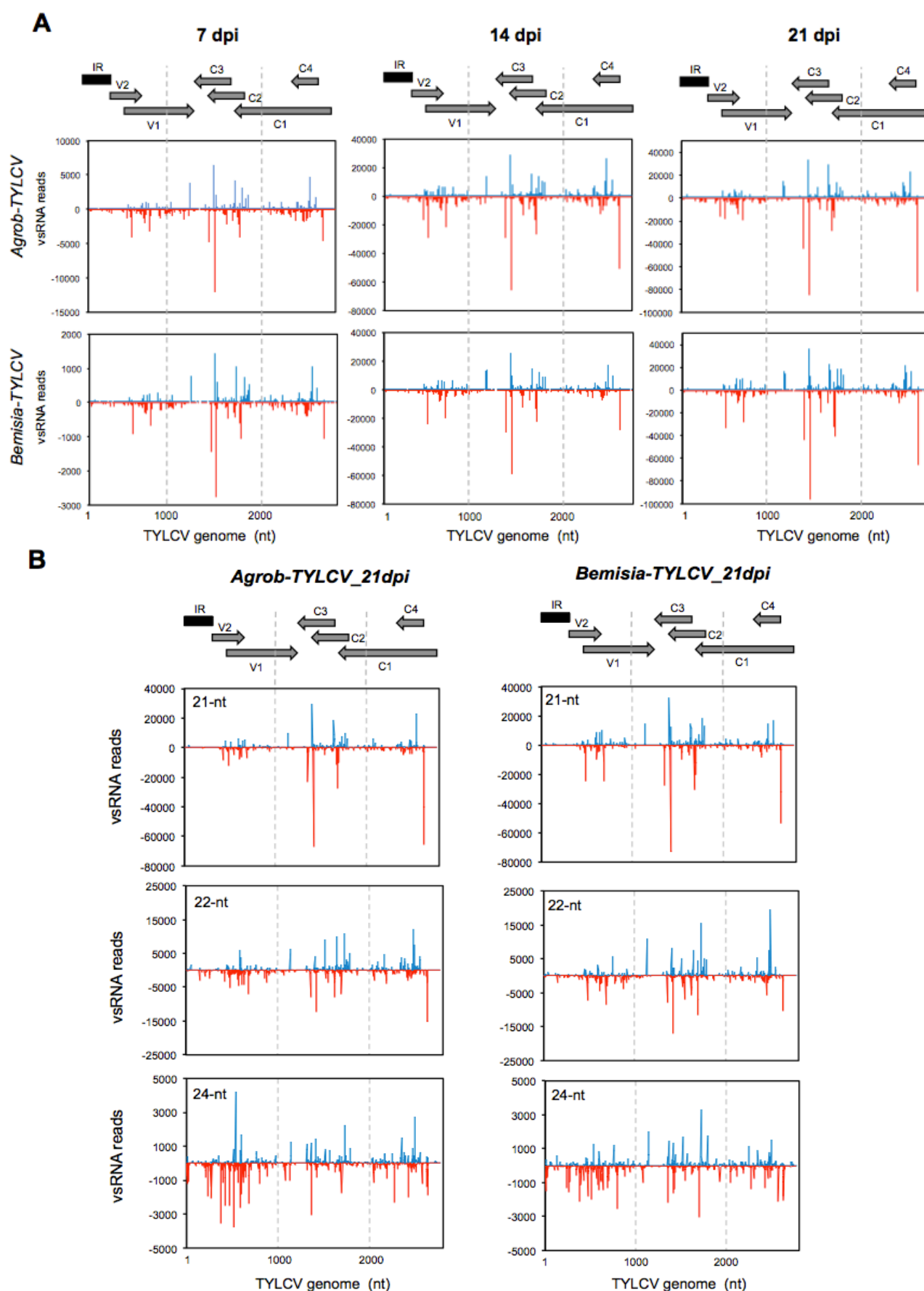
Figure 4. Maps from TYLCV vsRNAs. Tomato plants by either agroinoculation (*Agrob-TYLCV*) or using *B. tabaci* (*Bemisia-TYLCV*). Data from different time points during the systemic infection (7, 14 and 21 dpi) are shown. The

graphs plot the number of **(A)** 20-25 nt vsRNA or **(B)** 21, 22 and 24-nt vsRNA reads, at each nucleotide position of TYLCV genome (2781 nt). Bars above the axis (blue) represent sense reads starting at each respective position and those below (red) represent antisense reads ending at that position (please note the differences in the scale between 7dpi and the other two time points). One representative biological replicate is shown (replicate number is indicated in Table S2). The genome organization of TYLCV is shown schematically above the graphs. The predicted viral ORFs are shown as grey arrows and the IR as a dark rectangle.

Both sense and antisense vsRNAs displayed a heterogeneous distribution pattern along the genome with a large proportion of vsRNAs concentrating at specific areas, and this pattern was highly consistent between the two biological replicates at each time point assayed (Figure S1). Three regions corresponding to the overlapping (OL) sequences between V2/V1, C1/C2/C3 and C1/C4, named here as OL1, OL2 and OL3 respectively, contained more than 60% of the vsRNAs and showed the highest density of vsRNAs per nucleotide (Table 3).
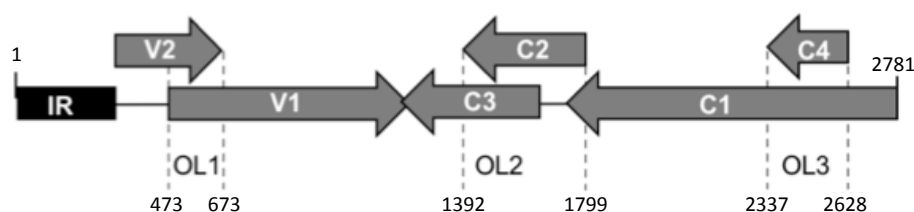
				Number of vsRNA per nucleotide						
				Agrob-TYLCV (dpi)			*Bemisia-TYLCV (dpi)*			
Region	**Start**	**End**	**Size**	**7**	**14**	**21**	**7**	**14**	**21**	
IR	1	313	313	30	207	232	6	114	192	
V2	314	472	159	76	739	615	5	413	543	
V2/V1	473	673	201	**343**	**1356**	**1336**	**72**	**1059**	**1430**	**OL1**
V1-C3	674	1391	718	119	569	481	75	476	587	
C1/C2/C3	1392	1799	408	**303**	**1627**	**1708**	**67**	**1317**	**2155**	**OL2**
C1- 3´end	1800	2336	537	109	534	488	23	361	473	
C1/C4	2337	2628	292	**255**	**1628**	**1741**	**55**	**1062**	**1616**	**OL3**
C1- 5´end	2629	2781	150	19	128	119	4	78	126	
Genome	1	2781	2781	146	832	818	31	616	893	

Table 3. Distribution of vsRNA abundance along TYLCV genome. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or by *B. tabaci* (*Bemisia-TYLCV*) and data from different time points during the systemic infection are shown (7, 14 and 21 dpi). The values represent the number of vsRNA (20-25 nt) per nucleotide at distinct genomic regions according to the position of the viral 5´nucleotide. The regions are identified by the viral genes they encompass, their size (nt) and their coordinates according to the TYLCV genome map are indicated at the top. The three genomic regions that showed the maximum number of vsRNA per nucleotide (OL1, OL2 and OL3) are marked in bold. Data comes from the same biological replicates showed in Figures 4A and 4B**.**

Some of the highest accumulated vsRNAs on those hotspots were strand specific (Figure 4A), suggesting the existence of potential strand-specific secondary structures used as DCL substrates. However, a good correlation between those vsRNAs and potential stem-loop

structures predicted by "mfold Web Server" (http://unafold.rna.albany.edu/?q=mfold/rna-folding-form) could not be identified. VsRNAs distribution in tomato plants infected with two TYLCV-related begomovirus described the existence of similar vsRNA hotspots at OL1 (TYLCSV and TYLCCNV) and OL3 (TYLCSV) but not at the OL2 region (Yang et al., 2011a; Miozzi et al., 2013). The vsRNA hotspots on both strands were interspersed with sequences that engendered lower vsRNA abundance, including: (i) the regions leftward and rightward of the transcription start sites of the V2 and C4 ORFs respectively, that encompass the IR, accumulated the lowest levels of vsRNAs (Table 3, positions 2629-313); (ii) the region in which the R and L transcriptional units are expected to overlap and potentially form a dsRNA precursor and (iii) the region located at the 3´-end of the C1 ORF, in front of the predicted transcription start site of C2/C3 mRNA that could correspond to a putative promoter region in TYLCV (positions 1852-1824).

Most of the vsRNA hotspots contained all the three major size-classes of TYLCV vsRNAs, 21, 22 and 24-nt, suggesting that the same dsRNA precursors may be processed by different DCLs (Figure 4B). However, a more careful analysis revealed that, while the distribution along the viral genome of 21 and 22-nt size classes was very similar, the 24-nt class showed a relative large accumulation at the IR at any time during the infection. The percentage of vsRNAs that were mapped to the IR was clearly higher for the 24-nt size class (from 7.4 to 9.1%) compared to the 21-nt (from 0.9 to 1.4%) or 22-nt (from 2.0 to 3.6%) classes (Table 4), indicating that the dsRNA substrate originated in that region is a preferential substrate for the DCL3 homologue which seems to be involved in the production of endogenous 24-nt sRNAs in tomato (Kravchik et al., 2014). Interestingly, the majority of the 24-nt vsRNAs that accumulated at the IR were strand-specific, with 3 to 5-fold more 24-nt vsRNAs mapping to the CS than to the VS strand (Table 4).

vsRNA at IR (%)

	Agrob-TYLCV						Bemisia-TYLCV					
	21-nt		22-nt		24-nt		21-nt		22-nt		24-nt	
	VS	CS	VS	CS	VS	CS	VS	CS	VS	CS	VS	CS
7 dpi	0.5	0.6	0.7	1.4	1.2	**6.4**	0.4	0.5	0.8	1.2	1.5	**7.5**
14 dpi	0.6	0.7	1.0	1.9	1.7	**5.7**	0.5	0.5	0.9	1.2	1.7	**5.5**
21dpi	0.6	0.8	1.0	2.6	1.5	**7.6**	0.6	0.5	1.1	1.5	2.3	**5.7**

Table 4. Percentage of 21, 22 and 24-nt vsRNAs classes at the IR from TYLCV. Tomato plants infected by either agroinoculation (*Agrob-TYLCV*) or by *B. tabaci* (*Bemisia-TYLCV*) and data from different time points during the systemic infection (7, 14 and 21 dpi) are shown. The percentage (%) represent the reads of each vsRNA size class from the IR (positions 1 to 313), at each strand (VS or CS) or from both strands (Total), relativized to the total TYLCV reads from that size. Data comes from the same biological replicates showed in Figures 4A and 4B**.**

In summary, this vsRNA distribution pattern along the TYLCV genome indicated that during infection, viral dsRNA precursors present preferential internal excisions of vsRNA at certain regions or they produce vsRNA duplexes with differential stabilities. Moreover, the promoter and terminator regions on the TYLCV genome are depleted of highly abundant vsRNAs.

No significant differences in the distribution of the vsRNA in the pool of 20-25 nt reads along the TYLCV genome, were observed among the samples collected at different times of the infection or between the plants treated by *Agrobacterium* or *Bemisia* (Figure 4, Table 3, Figure S1).
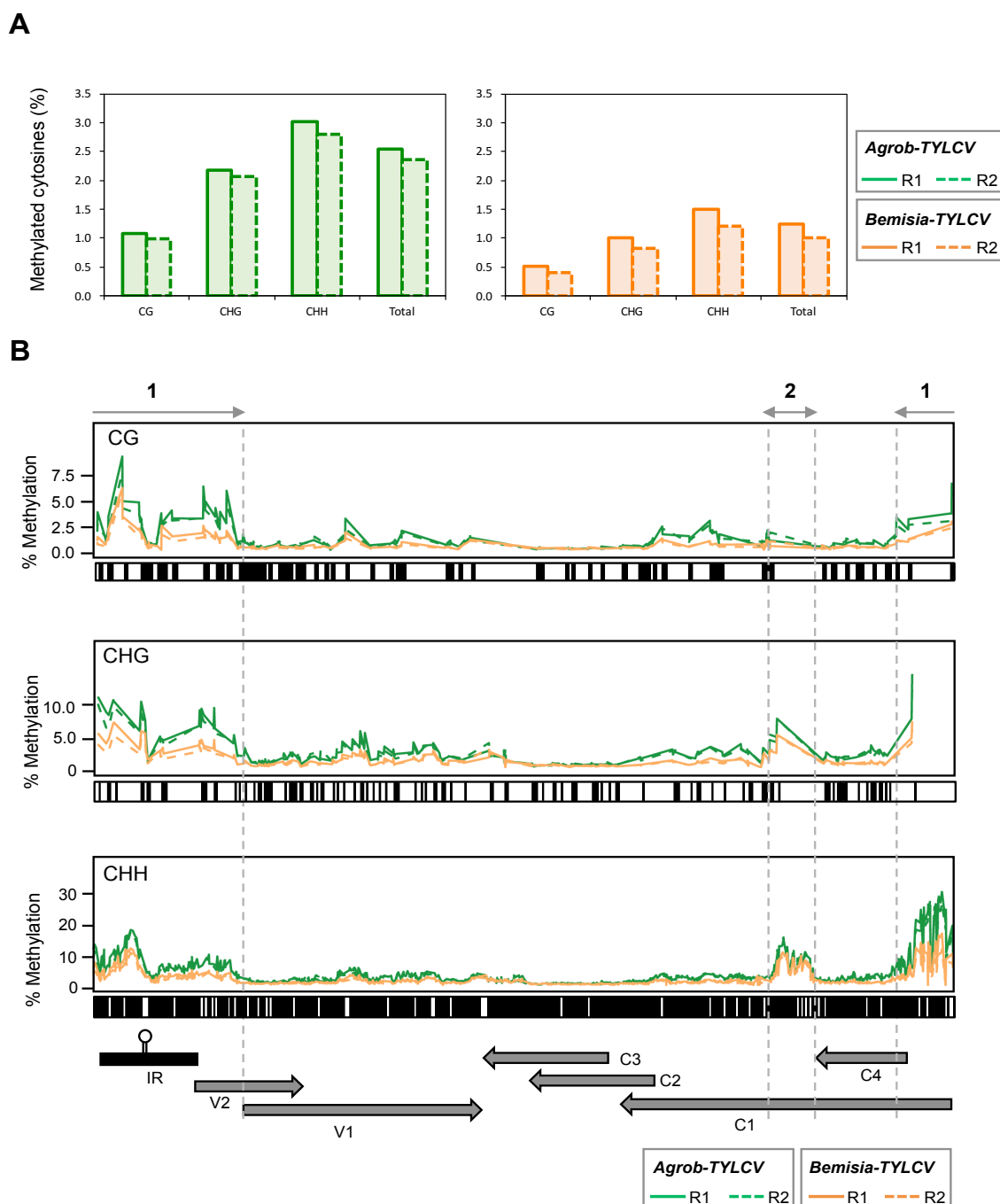
It has been reported that the 5′-terminal nucleotide of the sRNA directs its loading to different AGO complexes (Kim, 2008; Mi et al., 2008; Mallory and Vaucheret, 2010). The nucleotide of the 5′-terminal position was analysed in the vsRNAs from *Agrobacterium* and *Bemisia*-treated plants. We found a preference of beginning with a uridine (U) or an adenosine (A) in a similar percentage (from 30 to 40% for each of the two nucleotides) and a clear tendency to avoiding guanidine (G) in vsRNAs from 20 to 25-nt regardless of the infection method and time point assayed (Figure S2). In *Arabidopsis*, AGO2 and AGO4 preferentially recruit sRNAs with a 5′-terminal A, whereas AGO1 predominantly favours a 5′ terminal U (Kim, 2008; Mi et al., 2008; Mallory and Vaucheret, 2010). Our results suggest that TYLCV vsRNA might be loaded into diverse AGO-containing silencing complexes in tomato, as previously reported for TYLCCNV (Yang et al., 2011a). This scenario is similar to the one described for CaLCuV (*Cabbage leaf curl virus*) vsRNAs generated in *Arabidopsis*-infected plants, which were also enriched in 5′A and 5′U (70 to 60%) (Aregger et al., 2012).

1.4. High resolution methylome of TYLCV reveals dense methylation levels at two distinctive regions.

As previously mentioned, several reports have demonstrated that geminiviral DNA is methylated during infection and host recovery, and the level and extent of this methylation varies depending on the geminivirus and the host analysed. Nevertheless, they provide a slender picture of the geminiviral methylation pattern in infected plants, since the data is obtained by sequencing a limited number of RCA (Rolling Circle Amplification) or PCR-amplified clones, and in most cases, the level of methylation is determined only for a small region of the genome, generally the IR and flanking sequences. Here, deep sequencing of bisulfite-treated DNA (BS-Seq) was performed on the same two biological replicates of TYLCV-infected tomato plants used to determine the TYLCV transcriptome and vsRNA composition at 14 dpi. On average, 133 million 100 bp paired-end reads per sample (range, 128-144 million) were obtained (Table S2), which contained 65-107K reads uniquely mapped to the TYLCV genome, representing an average TYLCV methylome of 5,000x and 7,500x in plants treated with *Agrobacterium* or *Bemisia*, respectively (Table S7). Thus, to our knowledge, this work presents the highest-resolution methylome of a geminivirus and provides a reliable landscape for TYLCV methylation in tomato plants.

The percentage of methylated cytosines detected in TYLCV genome on each biological replicate ranged from 1.0% in *Bemisia*- to 2.5% in *Agrobacterium*-treated plants (Figure 5A), which were similar to the ones found in the methylome of TYLCCNV infecting *N. benthamiana* (Yang et al., 2011b). The average TYLCV methylome coverage was greater in *Bemisia*- than in *Agrobacterium*-treated plants (Table S7), ruling out the possibility that the lower levels for TYLCV methylation detected in the whitefly-treated plants could be due to technical problems. A

possible explanation for the differences detected in TYLCV DNA methylation between both methods could be the slight delay in the infection development observed at 7dpi in the plants TYLCV-infected by the whitefly (Figure 1, Figure 3, Table S3, Table S6). In spite of this difference, the relative proportion of methylated cytosines at the three contexts was similar by both infection methods, being the percentage of methylation at CHH and CG sites the highest and the lowest, respectively (Figure 5A, Table S8).

A

B

Figure 5. High resolution methylome of TYLCV. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*, green) or using the whitefly *B. tabaci* (*Bemisia-TYLCV*, orange). **(A)** The histogram shows the percentage of cytosine residues methylated at each cytosine context (CG, CHG and CHH) and the total methylation in TYLCV genome at 14 dpi. Each bar corresponds to an independent biological replicate (R1, solid line and R2, dotted line). **(B)** The graphs show the percentage of methylated cytosines in different contexts, CG, CHG and CHH, at each

position in TYLCV genome (please note the differences in the scale between the three cytosine contexts). Each line corresponds to an independent biological replicate (R1, solid line and R2, dashed line). The position of each cytosine context on TYLCV genome, is represented by a dark line underneath each graph. The genome organization of TYLCV is shown schematically underneath the graphs. The two highest methylated regions are indicated by a grey solid line with arrowheads (1 and 2).

Since the main target of geminiviral DNA methylation is dsDNA (Paprotka et al., 2011; Deuschle et al., 2016), this apparently low level of methylation has to be taken cautiously. A more precise value of the percentage of methylated geminiviral DNA could be estimated considering the percentage of cytosines that form part of the viral dsDNA molecules. At 14 dpi, an average of 11% and 18% of the total viral DNA strands form dsDNA molecules in *Agrobacterium* and *Bemisia*-treated plants, respectively (Table S8). Taking into account this rough calculation, we could estimate that the percentage of methylated cytosines in viral dsDNA is around 23% for the agroinfiltrated samples and 7% for the *Bemisia*-treated ones (Table S8). Moreover, a methylation value in a defined cytosine residue above 11% and 18% will indicate that, this residue will be methylated in all dsDNA viral molecules.

Interestingly, when checking cytosine methylation profiles across the viral genome, two distinctive regions with dense levels of cytosine methylation (marked as region 1 and 2) could be noticed in both infection methods, being highly consistent between the biological replicates (Figure 5B). As previously described for other geminiviruses, we detected high levels of methylation at the IR that spanned rightwards and leftwards and drastically dropped at the beginning of the V1 and C4 ORFs (region 1). Nevertheless, this high-resolution approach let us dissect the methylation pattern at this region, characterized by containing two divergent promoters. The lowest level of methylation at the IR was located at the 3′-end of the stem loop, in a sequence that contains the putative C2 protein binding domain (positions 195 to 203; CLE, conserved late element (Argüello-Astorga et al., 1994)) (Figure 6). At the 3′-side of the stem loop, in the region that comprises the TATA box of V2/V1 transcripts and the 5′-end of the V2 ORF, a moderate increase in DNA methylation was observed. This methylation dropped dramatically just a few nucleotides before the start site of V1 (Figure 6, region 1). At the 5'-side of the stem loop up to the C4-ORF start site, a larger rise in methylation levels was detected. This region, partially depleted of cytosines in the CG and CHG context, showed the highest level of DNA methylation in the viral genome (Figure 5B). Interestingly, a slight reduction in the methylation level of this region was observed before the start site of the C1 ORF, in a sequence that contains the iterons of the Rep binging domain and the TATA box of the L transcriptional unit (Figure 6). A second region (region 2), that spans approximately 150 nucleotides just downstream of the C4 ORF stop codon and that is also depleted from cytosines in CG or CHG context, contained elevated levels of cytosine methylation (positions 2182-2334) (Figure 6). This methylation peak seems to be upstream of the region that encompasses the formerly described elements that act as an internal promoter for C2 and C3 genes in different geminiviruses (Shivaprasad et al., 2005; Tu and Sunter, 2007; Borah et al., 2016) and the nucleosome-free gap described for *Abutilon mosaic virus* (AbMV) transcriptionally active minichromosomes (Deuschle et al., 2016).
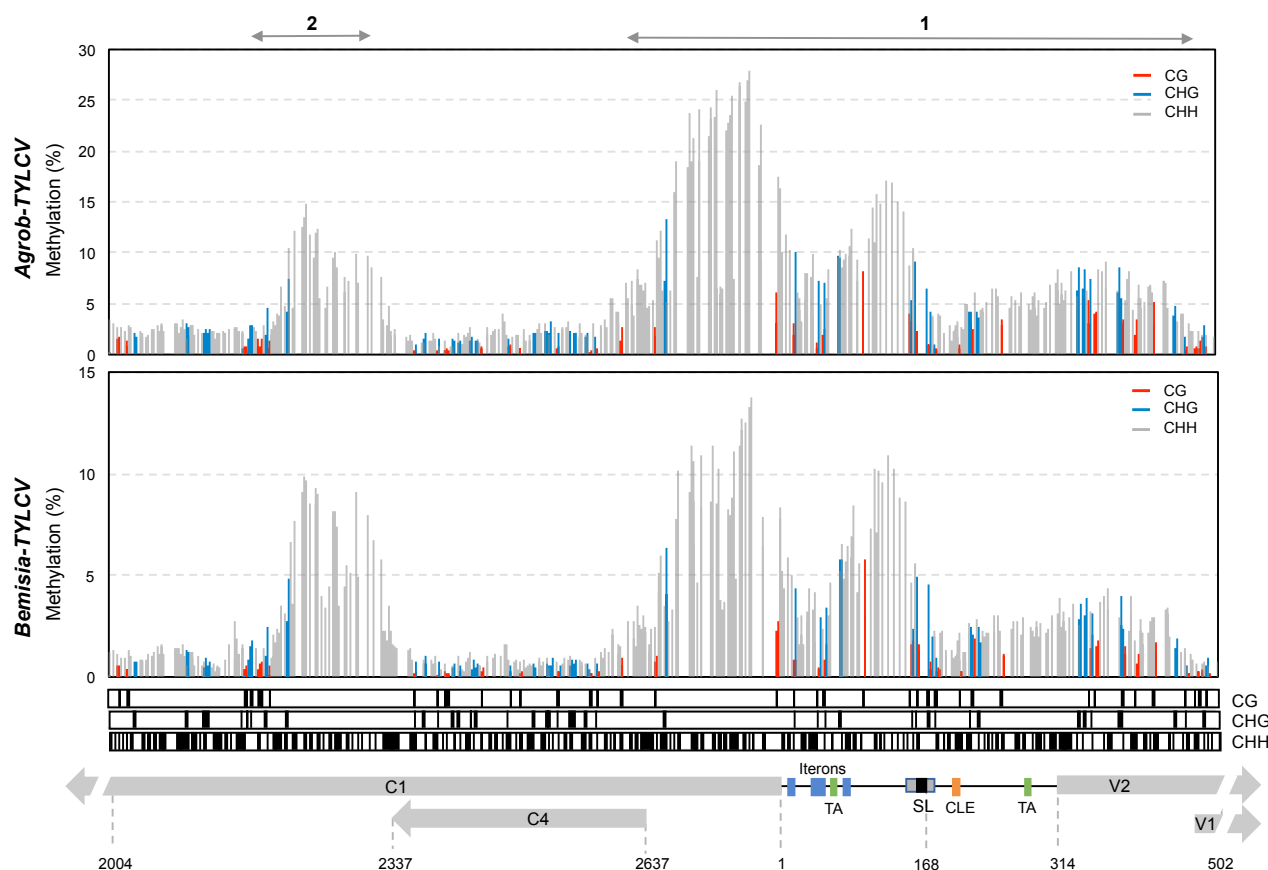
Figure 6. Detailed representation of TYLCV DNA methylation at the IR and flanking sequences. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or using the whitefly *B. tabaci* (*Bemisia-TYLCV*). The data correspond to the average of the two biological replicates shown in Figure 5. Bars represent the percentage of methylation on each cytosine context (CG in red, CHG, in blue and CHH in grey). The position of each cytosine context on TYLCV genome, is represented by a dark line underneath the graph. TYLCV genome organization at the IR and surrounding sequences is indicated at the bottom and the regulatory elements from the IR are shown (SL, stem loop; Iterons of the Rep-binding region; TA, TATA box; CLE, conserved late element). The predicted viral ORFs are shown as grey arrows and relevant nucleotide positions are indicated. The two highest methylated regions are indicated by a grey solid line with arrowheads (1 and 2).

DISCUSSION

Since the first report in 1988 (Elmer et al., 1988), agroinoculation has been widely used and accepted as a standard method to infect plant with gemiviruses. Although, along these years the scientific community has assumed that the dynamics of geminiviral infections produced by agroinoculation, or by its natural vector, the whitefly *B. tabaci*, were similar, this is the first time that a thorough molecular analysis of a side by side infection using both inoculation methods has been carried out. The data showed no significant differences in either of the parameters analysed (accumulation of viral DNA, viral transcripts, vsRNA and DNA methylation pattern) among the samples collected from plants infected by the two methods. This result confirms that

both are equivalent methods to study geminiviral infections.

24-nts long siRNAs, along with Pol IV, Pol V, RDR2, DCL3 and AGO4, are the essential players of the canonical RdDM pathway, but several RdDM mechanisms which partly incorporate components associated with PTGS such as 21 and 22-nt siRNAs, namely non-canonical, have also been reported (reviewed in Matzke and Mosher, 2014). The fact that elements of the RdDM canonical pathway seem to be involved in geminiviral genome methylation (Raja et al., 2008; 2014; Jackel et al., 2016), suggests that vsRNAs could be directing viral DNA methylation. To better appreciate the connection between TYLCV vsRNAs distribution, viral methylome and transcripts levels, data from TYLCV-agroinfiltrated plants at 14 dpi were compiled in Figure 7.
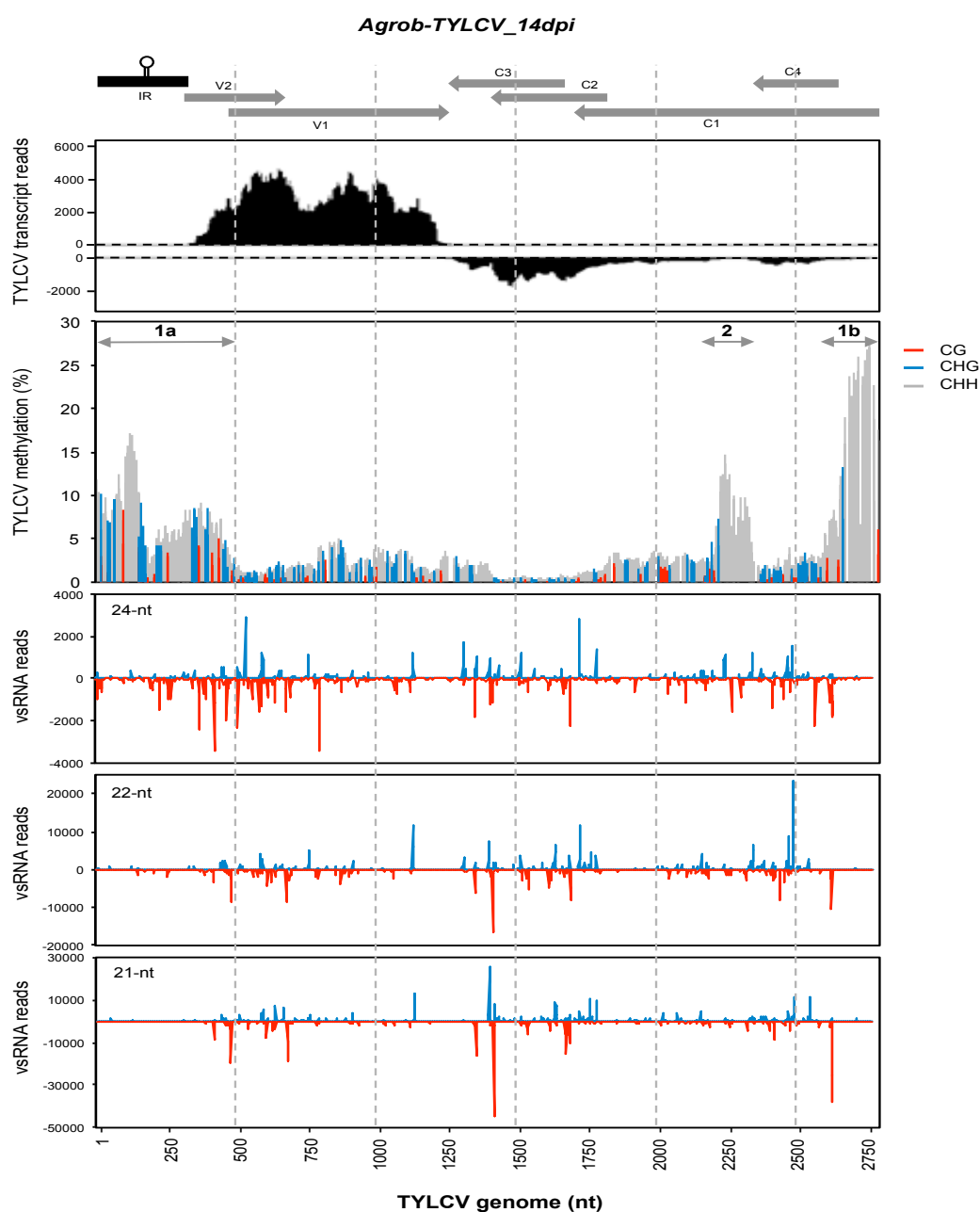

Figure 7. Maps of viral transcripts reads, DNA methylation and 24, 21 and 22-nt vsRNA reads, at each nucleotide position of TYLCV genome (2781 nt). Data from agroinfiltrated tomato plants (*Agrob-TYLCV*) at 14 dpi are shown. The genome organization of TYLCV is shown schematically above the graphs. Figure details are explained in Figure 2, 4B and 5B.

We could detect high methylation at sequences containing or adjacent to the three geminiviral promoters but, in general, the promoter and terminator regions were depleted of highly abundant vsRNAs. The low representation of vsRNAs that we have found at the IR of TYLCV is in accordance with the results previously obtained by high resolution RNA blots (Noris et al., 2004; Chellappan et al., 2004) or deep sequencing (Yang et al., 2011a; Aregger et al., 2012; Miozzi et al., 2013; Rogans et al., 2016) for other geminiviruses. Broadly, we did not find a clear correlation between 24, 21 and 22-nt vsRNA hotspots and viral DNA methylation peaks, as we could detect regions that contained a high density of 24-nt vsRNA and just residual cytosine methylation (Figure 7, positions 1390 to1811) as well as other areas that contained high levels of DNA methylation and low amounts of 24, 21 and 22-nt vsRNA (Figure 7, positions 2632 to 2871, 5′-end of C1). A detailed analysis of vsRNA distribution and DNA methylation at the promoters, described three different epigenetic scenarios. The region that contains the promoter for the R transcriptional unit showed elevated methylation levels at the three cytosine contexts and high density of 24-nts vsRNAs with limited presence of 21 and 22-nt vsRNAs, suggesting that cytosine methylation at this promoter region may be targeted by the canonical RdDM pathway (Figure 7, region 1a). This pattern was maintained through the V2 ORF up to the start site of V1, where we could detect an abrupt drop in DNA methylation and an increase in the accumulation of 21 and 22-nt vsRNAs, while the density of 24-nt size class was maintained (Figure 7, region 1a). A similar situation was observed at the L transcriptional unit promoter, with high levels of methylation and relative abundance of 24-nt vsRNAs, but with a low amount of the 21 and 22-nt classes (Figure 7, region 1a). DNA methylation also extended along the 5′end of the C1 ORF (position 2632 to 2781) dropping sharply at the beginning of C4 ORF. However, the epigenetic profile for the 5′end of the C1 ORF was different from the V2 one, as it also showed high levels of DNA methylation (the highest on TYLCV methylome) but accumulated almost no vsRNAs of any size, suggesting that DNA methylation at this region was RdDM-independent (Figure 7, region 1b). In plants, DNA methylation also occurs in a sRNA-independent manner and relies on chromatin features (Zemach et al., 2013; Stroud et al., 2014), opening up the possibility that histone modifications in the viral minichromosome at that region renders the DNA methylation at the 5′-end of the C1 ORF. Surprisingly, the chromatin in that region from another begomovirus does not contain H3K9me2 (Castillo-González et al., 2015), an epigenetic histone mark that could drive DNA methylation at the CHH context, opening the question about how DNA methylation is established in that region. We could detect a third epigenetic profile that corresponded to the methylation peak located around the middle of the C1 gene, just downstream the C4 ORF, which contained high levels of methylation, a high density of 24-nts vsRNAs from both strands and also a moderate amount of 21 and 22-nts vsRNAs (Figure 7, region 2). This observation suggested that canonical and non-canonical RdDM mechanisms could be responsible for establishing and maintaining the methylation at this enriched CHH region. Interestingly, moderate but dense levels of methylation were observed in almost all cytosines upstream the region that encompasses the formerly described elements that act as an internal promoter for C2 and C3 genes in different geminiviruses, which is activated by the Rep-mediated repression of its upstream promoter located at the IR (Shivaprasad et al., 2005; Tu and Sunter, 2007; Borah et al.,

2016). The higher accumulation of C2 and C3 transcripts reads from our transcriptomic data compared to C1 and C4 ones (Figure 7, Figure 2), supports the existence of an internal C2/C3 promoter also in TYLCV. The identification of the *cis* elements and the role that methylation could play in controlling its transcription, will need to be further characterized.

The type and amount of DNA methylation and its location within a gene (promoter or ORF) determines the impact of this epigenetic mark on its transcription. Our results indicated that in the viral population, both bidirectional promoters at the IR are methylated to a different extent, which could explain the differences found in the expression of the R and the L transcriptional units (Figure 7, Table 1). As we previously mentioned, it is accepted that V2 and V1 genes are expressed from the same transcriptional R unit, but our data showed that the RNA-Seq reads from both ORFs accumulated to different levels (Figure 2). The fact that a large region of the 5´-end of the V2 ORF (around 150 nt) showed high and dense DNA methylation levels (Figure 6, Figure 7), brings up the question whether that epigenetic mark could be involved in the lesser accumulation of V2 transcripts and favouring another transcription start site for V1. Interestingly, there is a TATA box (positions 448-451) 22-nt upstream of the V1 start codon (position 474) raising the possibility of an additional transcription start site that could produce specific transcripts for TYLCV V1. The same rationale could be applied to the C1 and C4 ORFs which share the same promoter at the IR. There is a minor accumulation of C1 transcripts compared to C4 in combination with high and dense levels of DNA methylation at its 5´-end just prior the start site of the C4 ORF.

The complex picture for the vsRNA profile and the viral DNA methylation landscape described here, poses the question whether the role of DNA methylation could be just narrowed down to a plant defence mechanism that transcriptionally silences the virus, or if the virus needs this epigenetic mark, and others, to properly regulate its gene expression. The mechanisms that lead to these distinct methylation/vsRNA accumulation patterns at the viral genome and the biological relevance of these epigenetic marks, will need further study and raise new exciting questions to the field.

Supplementary data.

Viral DNA	Agrob-TYLCV				Bemisia-TYLCV			
	2 dpi	7 dpi	14 dpi	21dpi	2 dpi	7 dpi	14 dpi	21dpi
VS ($\times 10^6$)	ndt	16±5.5	140±43	97±3.0	ndt	3.5±2.0	100±32	183±56
CS ($\times 10^6$)	ndt	1.1±0.4	6.9±1.5	4.3±0.7	ndt	0.22±0.1	6.9±0.9	9.0±0.5
VS/CS ratio	ndt	13.6	19.9	22.6	ndt	16.9	14.6	20.2

Table S1. Absolute quantification of virion-sense (VS) or complementary-sense (CS) strands of TYLCV during the infection. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or by *B. tabaci* (*Bemisia-TYLCV*). Data from different time points during the systemic infection are shown (2, 7, 14 and 21 dpi). Values represent the average ± SE of viral molecules ($\times 10^6$) per ng of DNA from the three biological replicates. The ratio VS/CS at each dpi, is indicated for each dpi. Ndt: non-detected.

	RNA-Seq: Raw read pairs (2x75-nt)				sRNA-Seq: Raw reads 16-50 nt (1x50-nt)			BS-Seq: Raw read pairs
	2 dpi	7 dpi	14dpi	21 dpi	7 dpi	14dpi	21 dpi	14dpi
Agrob-TYLCV_R1	28,308,689	34,325,652	31,509,779*	33,716,991	77,628,196	74,999,009^	75,720,534	137,255,859
Agrob-TYLCV_R2	32,494,007	36,180,219*	41,152,086	34,535,147*	82,137,526^	66,201,477	70,746,164^	129,137,303
Agrob-TYLCV_R3	32,770,672	27,175,266	30,003,595	30,713,610	nd	nd	nd	nd
Bemisia-TYLCV_R1	29,322,004	29,528,784	36,255,934*	25,898,937*	75,012,927	83,630,170^	79,764,671^	144,368,733
Bemisia-TYLCV_R2	29,357,804	36,254,389*	30,167,165	27,239,950	77,704,976^	81,084,958	82,169,778	128,106,473
Bemisia-TYLCV_R3	46,535,861	36,759,995	30,534,930	31,000,225	nd	nd	nd	nd

* Data obtained from this biological replicate is plotted on Figure 2
^ Data obtained from this biological replicate is plotted on Figure 4
nd: non-determined

Table S2. Raw reads for TYLCV-tomato infected samples from RNA-Seq, sRNA-Seq and BS-Seq analyses. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or using the whitefly *B. tabaci* (*Bemisia-TYLCV*) and data from different time points during the systemic infection (2, 7, 14 and 21 dpi) and from the different biological replicates (R1, R2 and R3) are shown.

	Reads mapped to TYLCV (No.)				Reads mapped to TYLCV (%)			
Agrob-TYLCV	*2 dpi*	*7 dpi*	*14dpi*	*21 dpi*	*2 dpi*	*7 dpi*	*14dpi*	*21 dpi*
R1	0	14477	86841	92538	0.000	0.022	0.110	0.141
R2	34	38532	57658	85343	0.000	0.056	0.101	0.132
R3	0	17785	61593	77002	0.000	0.035	0.103	0.130
Average					*0.00±0.000*	*0.04±0.017*	*0.10±0.005*	*0.13±0.006*
Bemisia-TYLCV	*2 dpi*	*7 dpi*	*14dpi*	*21 dpi*	*2 dpi*	*7 dpi*	*14dpi*	*21 dpi*
R1	3	5043	115250	95968	0.000	0.009	0.164	0.193
R2	10	26221	116090	91716	0.000	0.037	0.199	0.176
R3	10	6350	89807	106900	0.000	0.009	0.152	0.179
Average					*0.00±0.000*	*0.02±0.016*	*0.17±0.025*	*0.18±0.009*

Table S3. TYLCV transcripts reads from tomato plants infected by either agroinoculation (*Agrob-TYLCV*) or by *B. tabaci* (*Bemisia-TYLCV*). Data from each biological replicate (R1, R2 and R3) at different time points during the systemic infection are shown (2, 7, 14 and 21 dpi). Values correspond to the total number of reads that mapped to TYLCV genome (No.) or the percentage (%) that they represent relative to the total transcripts reads. The average ± SE from the three biological replicates is indicated for the percentage of TYLCV-mapped transcripts.

		Percentage (%) of each size-class of TYLCV vsRNA						
		20-nt	**21-nt**	**22-nt**	**23-nt**	**24-nt**	**25-nt**	**20-25 nt**
Agrob-TYLCV								
R1	*7 dpi*	0.01	0.10	0.05	0.01	0.02	0.00	0.20
R2	*7 dpi*	0.06	0.35	0.26	0.04	0.06	0.01	0.78
R1	*14 dpi*	0.40	2.11	1.50	0.25	0.41	0.05	4.72
R2	*14 dpi*	0.43	2.17	1.52	0.25	0.42	0.06	4.83
R1	*21 dpi*	0.52	2.38	1.70	0.32	0.43	0.07	5.40
R2	*21 dpi*	0.54	2.57	1.62	0.30	0.47	0.06	5.57
Bemisia-TYLCV								
R1	*7 dpi*	0.00	0.01	0.01	0.00	0.00	0.00	0.02
R2	*7 dpi*	0.01	0.08	0.06	0.01	0.01	0.00	0.17
R1	*14 dpi*	0.29	1.64	1.03	0.17	0.26	0.04	3.42
R2	*14 dpi*	0.24	1.38	0.93	0.15	0.23	0.04	2.96
R1	*21 dpi*	0.51	2.53	1.31	0.26	0.38	0.05	5.04
R2	*21 dpi*	0.54	2.63	1.41	0.28	0.37	0.06	5.28

Table S4. Size distribution of 20-25 nt TYLCV vsRNAs. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or using *B. tabaci* (*Bemisia-TYLCV*) and data from each biological replicate (R1 and R2) at different time points during the systemic infection are shown (7, 14 and 21 dpi). Values show the percentage (%) that each size-class of TYLCV vsRNA represent relativized to the total sRNA reads (19-35 nt).

	Reduntant vsRNA (%)					
	20-nt	**21-nt**	**22-nt**	**23-nt**	**24-nt**	**25-nt**
Agrob-TYLCV						
7 dpi	4.9	48.5	33.6	4.3	7.5	1.2
14 dpi	5.6	46.7	33.0	4.9	8.8	1.0
21 dpi	5.6	47.4	32.1	5.4	8.4	1.2
Bemisia-TYLCV						
7 dpi	4.9	49.4	32.9	4.5	6.9	1.4
14 dpi	5.5	49.6	31.8	4.5	7.5	1.1
21 dpi	6.3	53.0	27.7	4.8	7.2	1.1

	Unique vsRNA (%)					
	20-nt	**21-nt**	**22-nt**	**23-nt**	**24-nt**	**25-nt**
Agrob-TYLCV						
7 dpi	13.5	25.9	24.8	13.2	16.4	6.2
14 dpi	16.4	20.0	19.8	16.1	17.5	10.3
21 dpi	16.0	20.1	20.0	16.4	17.7	9.8
Bemisia-TYLCV						
7 dpi	10.8	32.2	28.3	10.8	13.7	4.1
14 dpi	16.4	20.4	20.2	15.9	17.2	9.9
21 dpi	16.6	20.0	19.5	16.2	17.2	10.4

Table S5. Percentage of redundant and unique 20-25 nt TYLCV vsRNAs from tomato plants infected by either agroinoculation (*Agrob-TYLCV*) or by *B. tabaci* (*Bemisia-TYLCV*). Values show the percentage (%) that each size-class of TYLCV vsRNA represent relativized to the total vsRNA reads (20-25 nt). The average from two independent biological replicates at different time points during the systemic infection (7, 14 and 21 dpi) are shown.

Redundant vsRNA

	20-nt		21-nt		22-nt		23-nt		24-nt		25-nt		20-25-nt	
	No.	% VS	No.	% VS	No.	% VS	No.	% VS	No.	% VS	No.	% VS	No.	% VS
Agrob-TYLCV														
7 dpi	12302	**50.7**	122459	**47.8**	84956	**51.1**	10866	**38.5**	19037	**39.6**	2931	**25.1**	252551	**47.8**
14 dpi	119774	**51.9**	1007651	**43.2**	711545	**48.7**	104932	**37.8**	190008	**40.3**	22520	**29.9**	2156433	**44.8**
21 dpi	132945	**51.3**	1123872	**41.9**	760109	**47.2**	127847	**36.4**	199502	**40.2**	27478	**26.8**	2371754	**43.5**
Bemisia-TYLCV														
7 dpi	1738	**50.9**	17451	**49.1**	11600	**51.7**	1578	**36.6**	2438	**34.2**	494	**22.3**	35300	**48.1**
14 dpi	89013	**50.1**	797537	**45.8**	510951	**50.3**	72861	**38.9**	120487	**39.6**	17853	**29.7**	1608700	**46.5**
21 dpi	161058	**52.5**	1355323	**42.8**	709965	**49.1**	123573	**38.1**	182897	**40.1**	26859	**28.7**	2559674	**44.6**

Unique vsRNA

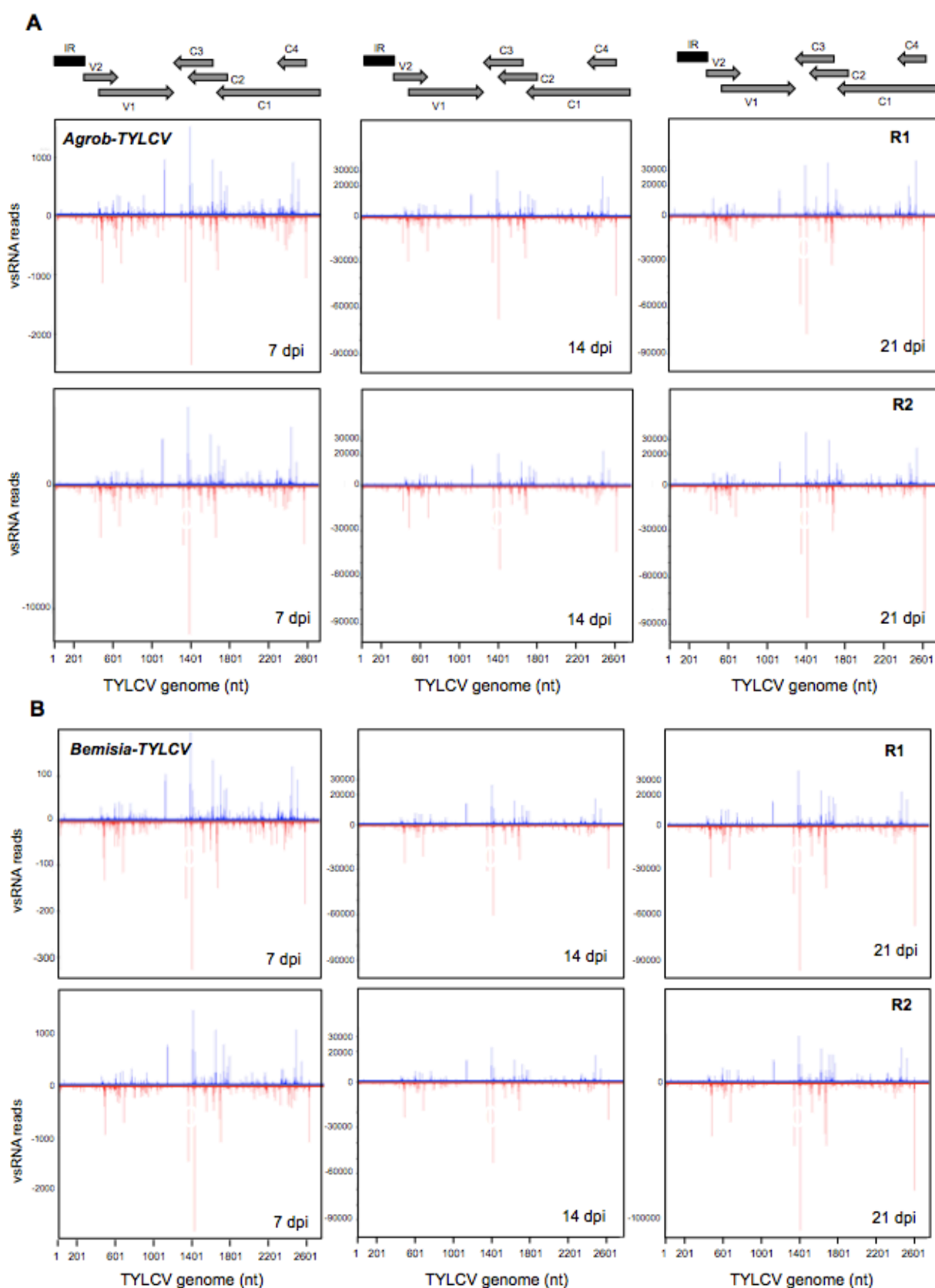|	20-nt		21-nt		22-nt		23-nt		24-nt		25-nt		20-25-nt	
	No.	% VS	No.	% VS	No.	% VS	No.	% VS	No.	% VS	No.	% VS	No.	% VS
Agrob-TYLCV														
7 dpi	1764	**49.3**	3385	**50.0**	3241	**50.0**	1719	**45.8**	2138	**46.2**	804	**39.8**	13050	**48.1**
14 dpi	3829	**50.2**	4684	**49.7**	4642	**49.9**	3763	**48.6**	4105	**48.9**	2401.5	**45.3**	23424	**49.0**
21 dpi	3653	**49.7**	4567	**49.4**	4553	**49.5**	3736	**48.8**	4024	**48.9**	2235.5	**46.2**	22768	**49.0**
Bemisia-TYLCV														
7 dpi	532	**48.6**	1582	**49.7**	1393	**50.6**	530	**58.5**	673	**58.3**	201	**64.8**	4911	**47.1**
14 dpi	3415	**49.8**	4253	**50.3**	4194	**50.3**	3316	**51.3**	3580	**51.1**	2056	**54.3**	20813	**49.0**
21 dpi	3626	**50.3**	4360	**50.5**	4263	**50.4**	3540	**51.4**	3756	**51.1**	2269.5	**53.1**	21813	**49.1**

Table S6. Total number (No.) and percentage mapped to the virion-sense strand (% VS) of TYLCV 20-25 nt vsRNA. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or using *B. tabaci* (*Bemisia-TYLCV*) and the average value from two independent biological replicates at different time points the during systemic infection (7, 14 and 21 dpi) are shown.

	Uniquely mapped read pairs	*Average coverage TYLCV*	*Average coverage TYLCV-mean*
Agrob-TYLCV_R1	78932	5588	5096
Agrob-TYLCV_R2	64817	4604	
Bemisia-TYLCV_R1	107547	7626	7585
Bemisia-TYLCV_R2	106295	7544	

Table S7. Average methylome coverage of TYLCV genome at 14 dpi. Tomato plants were infected by either agroinoculation (*Agrob-TYLCV*) or using the whitefly *B. tabaci* (*Bemisia-TYLCV*) and data from the two biological replicates (R1 and R2) are shown.

	Experimental data			Estimated data	
	Total meC	Total unmeC	Methylation (%)	meC-dsDNA	Methyl.-dsDNA (%)
Agrob-TYLCV_R1 (14.1% strands in dsDNA)					
CG	5,147	467,270	1.09	66,611	7.7
CHG	11,485	517,289	2.17	74,557	15.4
CHH	56,788	1,830,324	3.01	266,083	21.3
Total	*73,420*	*2,814,883*	*2.54*	*407,251*	***18.0***
Agrob-TYLCV_R2 (8.6% strands in dsDNA)					
CG	3,895	385,954	1.00	33,527	11.6
CHG	8,992	426,385	2.07	37,442	24.0
CHH	42,768	1,487,033	2.80	131,563	32.5
Total	*55,655*	*2,299,372*	*2.36*	*202,532*	***27.5***
Bemisia-TYLCV_R1 (13.5% strands in dsDNA)					
CG	3,286	638,877	0.51	86,692	3.8
CHG	7,383	723,674	1.01	98,693	7.5
CHH	38,251	2,506,805	1.50	343,583	11.1
Total	*48,920*	*3,869,356*	*1.25*	*528,967*	***9.2***
Bemisia-TYLCV_R2 (22.5% strands in dsDNA)					
CG	2,671	642,910	0.41	145,256	1.8
CHG	6,037	728,050	0.82	165,170	3.7
CHH	30,588	2,509,445	1.20	571,507	5.4
Total	*39,296*	*3,880,405*	*1.00*	*881,933*	***4.5***

Table S8. Total number of methylated (meC) and unmethylated cytosines (unmeC) and the percentage of methylation (%) of TYLCV genome at 14 dpi obtained from tomato plants infected by agroinoculation (*Agrob-TYLCV*) or using *B. tabaci* (*Bemisia-TYLCV*). The data from the two biological replicates (R1 and R2) are shown (Experimental data). The number of methylated cytosines present in dsDNA strands (meC-dsDNA) shown in "Estimated data" was calculated according to the percentage of strands in dsDNA form from each biological replicate (% underneath each replica name, with a mean value of 11% for TYLCV-agroinfiltrated plants and 18% for *Bemisia*-treated plants). These percentages were obtained from the quantification of CS and VS strands (Figure 1), considering that all CS strands form dsRNA (Rodriguez-Negrete et al., 2014). The estimated percentage of methylation (%) of dsDNA is also indicated.

Figure S1. Maps of vsRNA from TYLCV-infected tomato plants by either **(A)** agroinoculation (Agrob-TYLCV) or using **(B)** the whitefly *B. tabaci* (Bemisia-TYLCV). Data from both biological replicates for each condition (R1 and R2) at different time points during the infection (7, 14 and 21 dpi) are shown. Data from different time points during the systemic infection are shown (7, 14 and 21 dpi). The graphs plot the number of 20-25 nt vsRNA at each nucleotide position of TYLCV genome (2781 nt). Bars above the axis (blue) represent sense reads starting at each respective

position and those below (red) represent antisense reads ending at the respective position. The genome organization of TYLCV is shown schematically above the graphs and the rightward ORFs (C1, C4, C2 and C3), the leftward ORFs (V1 and V2) and the intergenic region (IR) are indicated. The predicted mRNAs are shown as grey arrows and the IR as a dark rectangle.
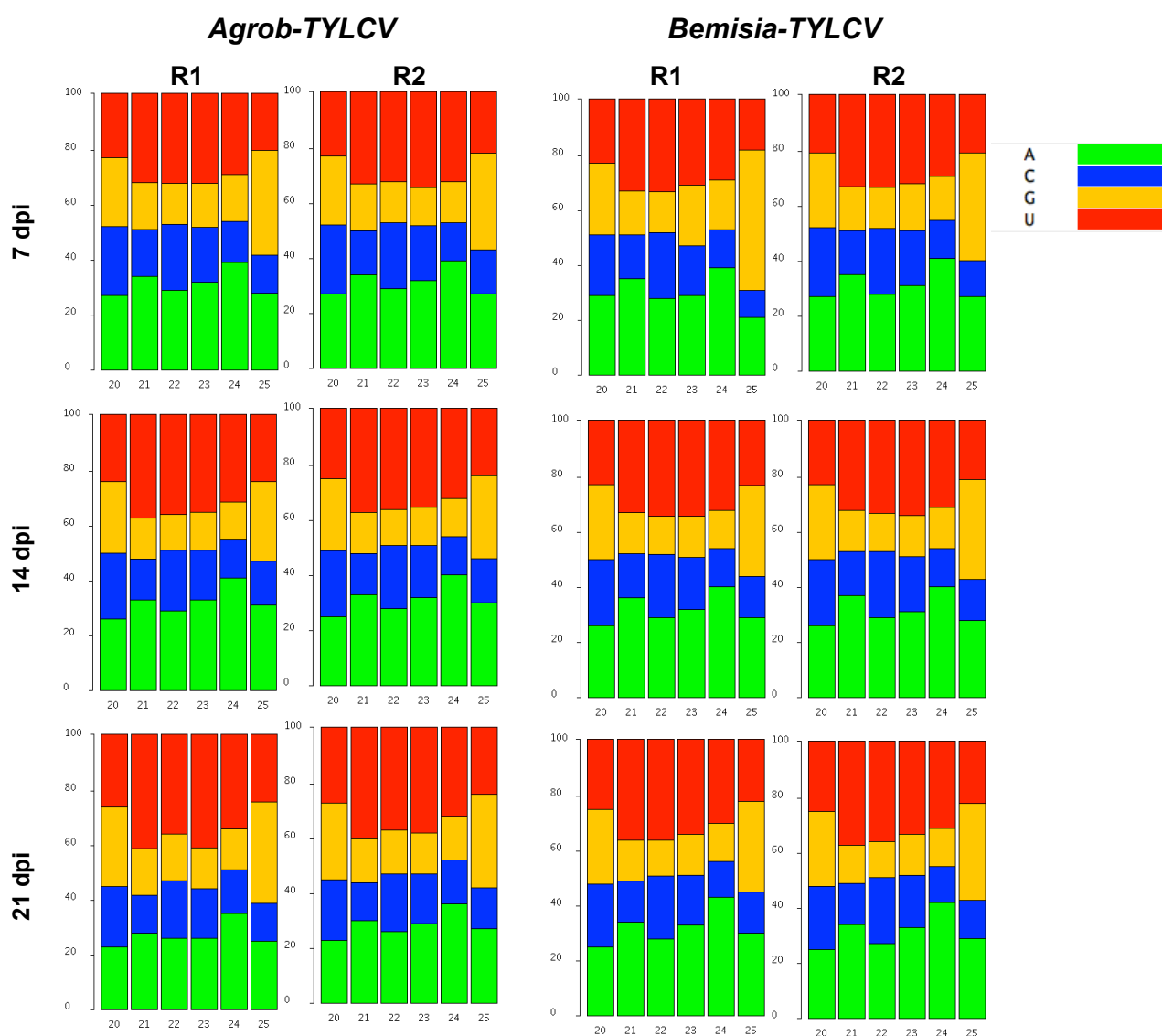
Figure S2. Abundance of vsRNAs 5'-terminal nucleotide with respect to the amount of individual nucleotides in the TYLCV genome. Tomato plants were infected by either agroinoculation (Agrob-TYLCV) or using the whitefly *B. tabaci* (Bemisia-TYLCV) and both biological replicates for each condition (R1 and R2) at different time points during the infection (7, 14 and 21 dpi) are shown. The Y-axis represents the percentage (%) of each nucleotide with respect to the total amount of that nucleotide. Both biological replicates are shown for each experiment (R1 and R2).

Chapter 2

Role of the plant DNA methylation machinery during geminivirus infection

BACKGROUND

DNA methylation has been extensively proposed as one of the main defence mechanism against geminivirus (Ermak et al., 1993; Wang et al., 2003, Raja et al., 2008; Buchmann et al., 2009; Rodríguez-Negrete et al., 2009; Yang et al., 2011; Zhang et al., 2011; Rodríguez-Negrete et al., 2013; Raja et al., 2010; Yang et al., 2013; Raja et al., 2014; Wang et al., 2014; Castillo-González et al., 2015; Sun et al., 2015; Ceniceros-Ojeda et al., 2016; Jackel et al., 2016; Coursey et al., 2018). Raja and colleagues (Raja et al., 2008) showed that methylation-deficient *A. thaliana* mutants in DNA methyltransferases (*met1-7*, *cmt3-7* and the double mutant *drm1-2 drm2-2*) were hypersusceptible to the infection by the curtovirus *Beet curly top virus* (BCTV) or the bipartite begomovirus *Cabbage leaf curl virus* (CaLCuV), since the plants displayed more severe symptoms (stunting and inflorescence deformation) when compared to wild-type infected plants. Similar results were reported for other DNA methylation-related mutants, such as the mutant in the chromatin remodeler *DDM1*, and mutants of genes involved in the RdDM pathway (*ARGONAUTE 4, DICER-LIKE 3* and a subunit of *DNA-DEPENDENT RNA POLYMERASE IV*) or in the methyl cycle (*ADK1* and *ADK2*). Conversely, the authors did not detect an increase of the viral titers in any of these mutants by Southern blot, but found slightly reduced levels of DNA methylation at the intergenic region (IR) of BCTV and CaLCuV in some of the mutant backgrounds. Additionally, a study from Castillo-González and collaborators (2015) demonstrated that an *A. thaliana* mutant in the H3K9 histone methyltransferase gene, *KRYPTONITE* (*KYP*), supported greater CaLCuV titers and increased susceptibility. A similar observation was also reported from a bipartite begomovirus, *Indian cassava mosaic virus* (ICMV), that accumulated to higher levels in *N. benthamiana* plants that displayed down-regulated transcript levels of *CMT3* and *KYP* (Sun et al., 2015). These findings suggested that DNA methylation and transcriptional gene silencing may be playing a role in defence against geminivirus.

To gain insight into the biological importance of DNA methylation during geminivirus infection, we have infected, with the monopartite begomovirus *Tomato yellow leaf curl virus* (TYLCV), *A. thaliana* plants containing mutations in genes that encode for DNA methyltransferases or a demethylase. Besides, we also carried out similar infections experiments in *N. benthamiana* plants impaired in the expression of the plant DNA methylation machinery.

RESULTS

2.1. Geminivirus infection of *A. thaliana* mutants involved in DNA methylation.

To characterize the biological relevance of the host DNA methylation machinery in TYLCV infection, we assessed the viral accumulation in *A. thaliana* plants defective in DNA

methyltransferases activities: the *met1-3* single mutant (*MET1* null allele) and the *drm1-2 drm2-2 cmt3-11* triple mutant (hereafter named *ddc*; null alleles for the three genes *DRM1*, *DRM2* and *CMT3*). In addition, we included in the analysis a mutant in *REPRESSOR OF SILENCING 1 (ROS1)* (*ros1-4* null mutant), a DNA glycosylase involved in active DNA demethylation.

2.1.1. TYLCV infection analysis in *met1-3*, *ros1-4* and *dc* plants.

Seeds from the three mutants were obtained from NASC (Nottingham Arabidopsis Stock Centre). Before infection, mutant seeds were germinated in the media with the appropriate antibiotic or in substrate to genotype them (Supplementary data from this Chapter). After genotyping the mutants, we concluded that the *ddc* seeds correspond to a homozygous double mutant *drm2-2 cmt3-11* (named *dc*), and not the triple mutant *drm1-2 drm2-2 cmt3-11* (*ddc*).

To perform the infection experiments, four-week-old wild-type and mutant plants were agroinoculated with TYLCV. Samples were collected at 21 days post-inoculation (dpi), the total DNA was extracted from the whole plant and the amount of virus was quantified by qPCR. First, 16 heterozygous *met1-3* mutant and 14 Col-0 (control) plants were infected in three independent biological replicates (R1, R2 and R3). The heterozygous *met1-3* mutant was previously described to induce DNA hypomethylation, to reactivate transcriptionally silent *loci* and to induce epigenetically-controlled phenotypes such as late flowering (Saze et al., 2003), suggesting that the lack of one of the *MET1* alleles is enough to revert transcriptional gene silencing. The mean values of TYLCV accumulation are represented in Figure 1. No statistically significant differences in the amount of TYLCV were observed in *met1-3* mutants compared to control plants (Col-0) (Figure 1).
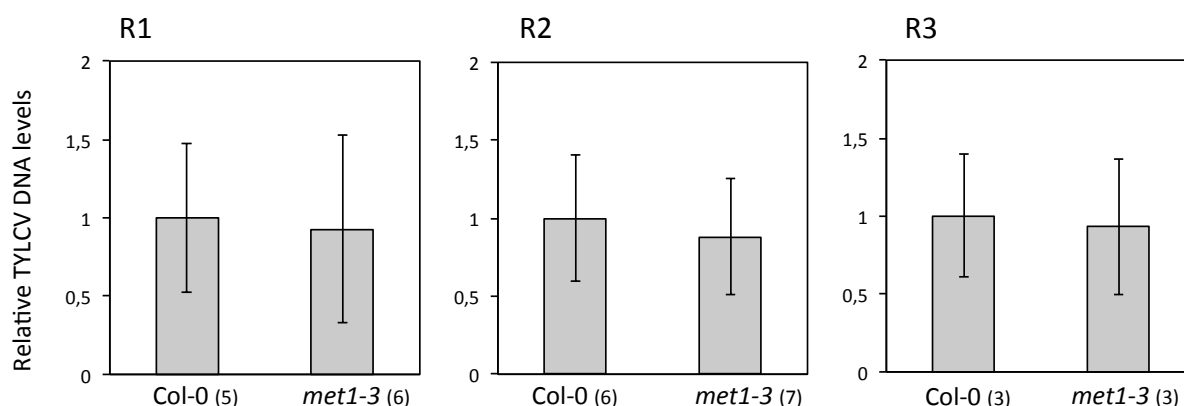

Figure 1. Relative TYLCV DNA levels determined by qPCR in *A. thaliana* Col-0 and *met1-3* plants. Four-week-old Col-0 and *met1-3* plants were agroinoculated with TYLCV. At 21 dpi, total DNA was extracted from the whole plant and TYLCV DNA levels were measured by qPCR. Each graph corresponds to one biological replicate (R1, R2 and R3). TYLCV DNA levels were normalized to *ACT2* and are presented as the relative amount of viral DNA compared with the amount found in Col-0 samples (set to 1)*. Bars represent mean values ± standard deviation. The number of plants used is shown (n). Statistics established no significant differences when compared Col-0 and *met1-3* plants, as determined by Student's t-test.

Next, we infected the homozygous *ros1-4* mutant and the homozygous *dc (drm2-2 cmt3-11)* double mutant and determined the TYLCV DNA levels from three independent biological replicates (R1, R2 and R3). The results showed similar viral levels in all of them (Figure 2).
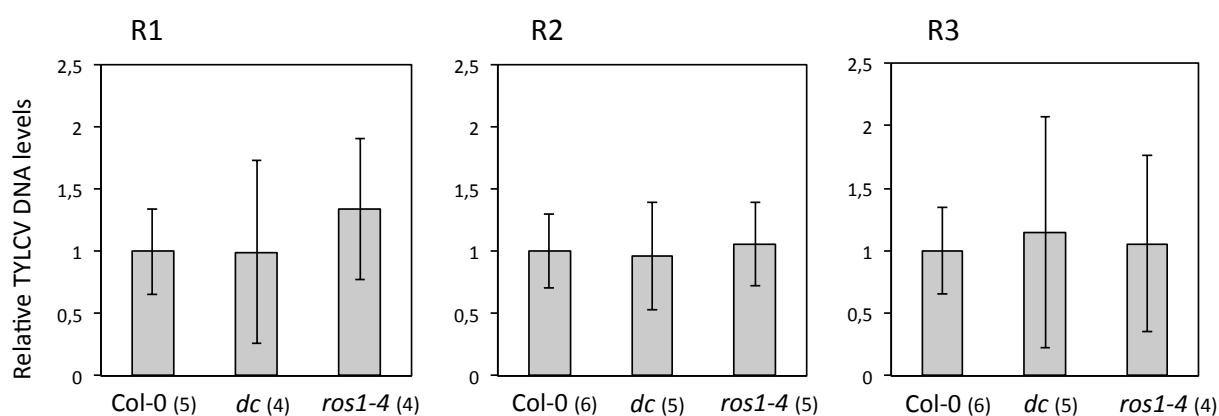

Figure 2. Relative TYLCV DNA levels determined by qPCR in *A. thaliana* Col-0, *dc* and *ros1-4* plants. Four-week-old Col-0, *dc* and *ros1-4* plants were agroinoculated with TYLCV. At 21 dpi, total DNA was extracted from the whole plant and viral DNA levels were measured by qPCR. Each graph corresponds to one biological replicate (R1, R2 and R3). TYLCV DNA levels were normalized to *ACT2* and are presented as the relative amount of viral DNA compared with the amount found in Col-0 samples (set to 1). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Statistics established no significant differences when compared Col-0 with both mutants *dc* and *ros1-4*, as determined by Student's t-test.

In summary, our data indicated that the impairment of the plant DNA (de)methylation machinery has no effect on TYLCV accumulation in *A. thaliana* (Figure 1 and 2). This observation contradicts previous results from Raja et al (2008), which show that methylation-deficient *A. thaliana* mutants (such as *met1-7, cmt3-7* or *drm1-2 drm2-2*) are hypersusceptible to geminivirus infection.

2.1.2. TYLCV and BCTV infection analysis in *ddc* plants.

To further characterize the effect of DNA methylation on geminiviral infection, we ordered a new set of *ddc* seeds from NASC, which were genotyped (please see details in Supplementary Data from this Chapter) and infected with TYLCV and BCTV. We included this latter virus in the analysis as it was previously shown that the *met1-7* and *cmt3-7* single mutants and the *drm1-1 drm2-2* double mutant were hypersusceptible to BCTV infection as compared with wild-type plants (Raja et el., 2008).

Data from our group (see Chapter 4) suggested that in *A. thaliana,* BCTV replicated to a different extent in floral tissue and the rosette leaves. As this geminivirus (and also TYLCV) is restricted to phloematic cells, one will expect to have a higher accumulation of this virus in flowers than in leaves, as inflorescences contain a higher proportion of vascular tissue. To determine if that entail a larger accumulation of viral DNA in the inflorescences than in the apical

rosette leaves, we quantified the amount of viral DNA in Col-0 plants infected with BCTV (10 plants) or TYLCV (7 plants), in three inflorescences and three apical leaves from each plant, at 26 dpi (Figure 3). The results showed that both viruses accumulated to higher levels in inflorescences than in leaves, and that effect was more pronounced in the plants infected with BCTV.
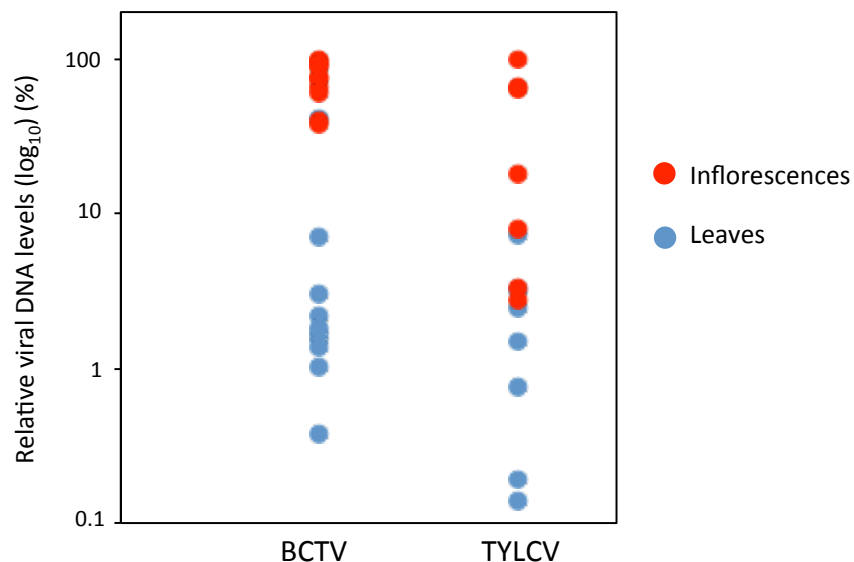

Figure 3. Relative TYLCV and BCTV DNA levels determined by qPCR in *A. thaliana* Col-0. Four-week-old Col-0 plants were agroinfected with TYLCV or BCTV. At 26 dpi, total DNA was extracted from apical leaves and inflorescences of each plant and DNA viral levels were measured by qPCR. Viral DNA levels were normalized to *ACT2* and are presented as the relative amount compared with the highest value for each virus (set to 100%). Each dot represents a sample.

According to this result, we determined the effect of the *ddc* triple mutant on a viral infection by analysing, separately, the inflorescences and the leaves from TYLCV- and BCTV-infected plants. DNA was extracted from three apical rosette leaves and three inflorescences from the same plant at 26 dpi, and the relative TYLCV and BCTV DNA levels were measured by qPCR. We observed that TYLCV DNA levels were not significantly different when comparing leaves or inflorescences between Col-0 and *ddc* plants (Figure 4A). The mean value of BCTV DNA levels resulted markedly higher in the leaves of *ddc* plants as compared with Col-0 leaves (more than seven times), although this difference in viral accumulation was not significant due to the variability of the infection between plants (Figure 4B). On the other hand, the low variability for viral DNA accumulation shown in floral tissue resulted in statistically significant higher BCTV amount in *ddc*, compared with Col-0 plants (Figure 4B).
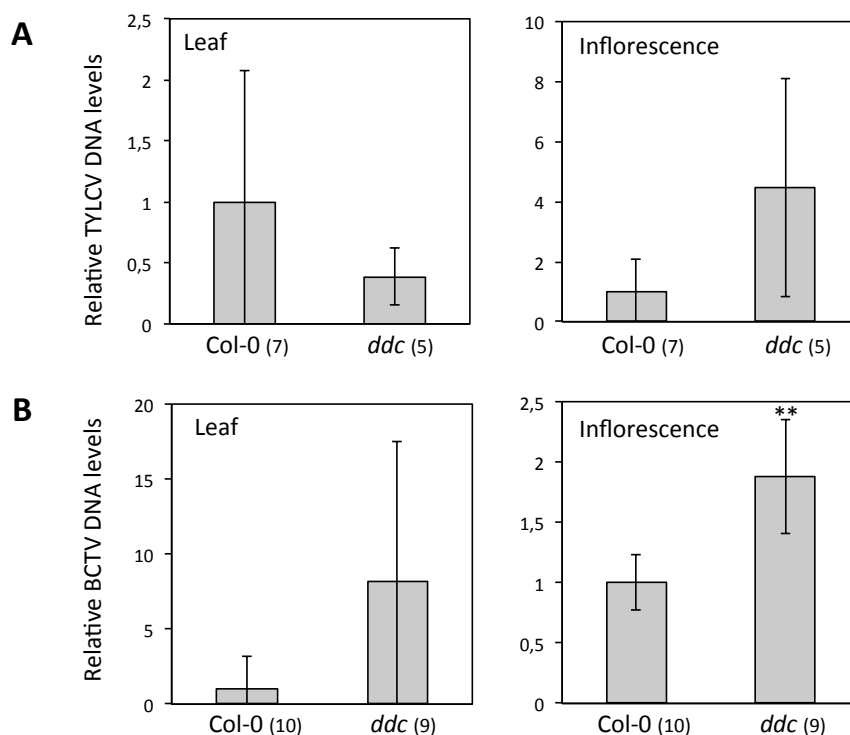
Figure 4. Relative TYLCV and BCTV DNA levels determined by qPCR in *A. thaliana* Col-0 and *ddc* plants. Four-week-old Col-0 and *ddc* plants were agroinoculated with TYLCV or BCTV. At 26 dpi, total DNA was extracted from three pooled apical leaves and three pooled inflorescences of each Col-0 and *ddc* plant, and **(A)** TYLCV and **(B)** BCTV DNA levels were measured by qPCR. Viral DNA levels were normalized to *ACT2* and are presented as the relative amount of viral DNA compared with the amount found in Col-0 samples (set to 1). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Asterisks indicate significant differences according to Student's t-test (**, p-value < 0.01).

Together, these results suggested that the deficiency in the maintenance methylation machinery at CG and CHG sites, conducted by MET1 and CMT3, did not affect TYLCV accumulation levels in *A. thaliana* plants (Figure 1 and 2). Besides, the impairment of one of the key enzymes involved in DNA demethylation, ROS1, also had no impact on TYLCV accumulation levels (Figure 2). Moreover, TYLCV levels were not significantly different in the leaves and inflorescences of wild-type and the *ddc* triple mutant, which is affected in the maintenance of DNA methylation at CHG and CHH sites and the establishment of *de novo* methylation at the three cytosine contexts (Figure 4). However, there was a greater accumulation of BCTV in *ddc* inflorescences (and also in leaves although with a high variability) compared to wild-type (Figure 4), which suggests that the plant DNA methylation machinery could be playing a role in defence against, at least, some geminivirus genera.

2.2. Geminivirus infection in DNA methylation-deficient *N. benthamiana* plants.

In order to analyse the biological importance of the host DNA methylation machinery on geminivirus infection in other plant hosts, we induced post-transcriptional silencing of key genes

involved in DNA methylation in *N. benthamiana* using VIGS (Virus-induced gene silencing). VIGS methodology uses a viral vector carrying a plant target gene sequence that is silenced as the virus-derived vector moves through the plant and synthesizes viral double-stranded RNAs, and thus, activates the RNA silencing machinery. We used the VIGS system based on *Tobacco rattle virus* (TRV) (Ratcliff et al., 2001) to evaluate the impact of the silencing of *NbMET1*, *NbCMT3*, *NbDRM2* and *NbROS1* on TYLCV infection.

Partial cDNA clones of *NbMET1, NbCMT3, NbDRM2* and *NbROS1* were cloned into the RNA2-based vector, pTV00 (Ratcliff et al., 2001, see Materials and methods)*.* To evaluate the silencing efficiency of each of the constructs, two fully expanded leaves of six-week-old *N. benthamiana* plants were co-infiltrated with two cultures of *Agrobacterium tumefaciens:* one carrying the clone for TRV RNA1 and another one carrying the TRV RNA2-based construct, named hereafter TRV-*NbMET1*, TRV-*NbCMT3*, TRV-*NbDRM2* or TRV-*NbROS1*. In addition, the TRV RNA2-based vector was also agroinfiltrated as a negative control (named TRV), and naïve plants were included in the analysis to rule out the possibility that TRV infection could alter the expression of the analysed genes. A TRV RNA2-based construct containing the phytoene desaturase (*PDS*) gene (TRV-*NbPDS*) was agroinfiltrated as a positive control. This gene is essential for the production of carotenoids, and the tissue in which *PDS* is silenced turns white due to photo-bleaching (Kumagai et al., 1995; Ruíz et al., 1998; Ratcliff et al., 2001). In our conditions, apical leaves from *N. benthamiana* plants agroinfiltrated with TRV-*NbPDS* showed a white-yellowish phenotype at 7-8 dpi (data not shown). At 18 dpi, total RNA was isolated from the second and third most apical leaves and the relative transcript levels of *NbMET1, NbCMT3, NbROS1* and *NbDRM2* were determined by RT-qPCR (Figure 5). The transcript levels of *NbEF1α* (*Elongation factor 1 α*) were also measured and used to normalize the transcript levels in each sample (Rotenberg et al., 2006). The plants agroinfiltrated with TRV-*NbMET1,* TRV-*NbCMT3* and TRV-*NbROS1* displayed a statistically significant transcriptional repression of *NbMET1* (87%), *NbCMT3* (87%) and *NbROS1* (62%), respectively, when compared with TRV plants (Figure 5A). However, we could not detect the down-regulation of *NbDRM2* in plants agroinfiltrated with TRV-*NbDRM2* (data not shown). The relative transcript levels of *NbMET1, NbCMT3* or *NbROS1* were similar between control plants (just agroinfiltrated with TRV) and naïve plants, confirming that TRV infection does not alter significantly the expression of these three genes (Figure 5A). Finally, we measured the relative TRV transcript levels in all the assayed plants to determine if silencing of these genes could impair TRV accumulation. As shown in Figure 5B, no significant differences were found between TRV plants and those in which *NbMET1, NbCMT3* or *NbROS1* were silenced.
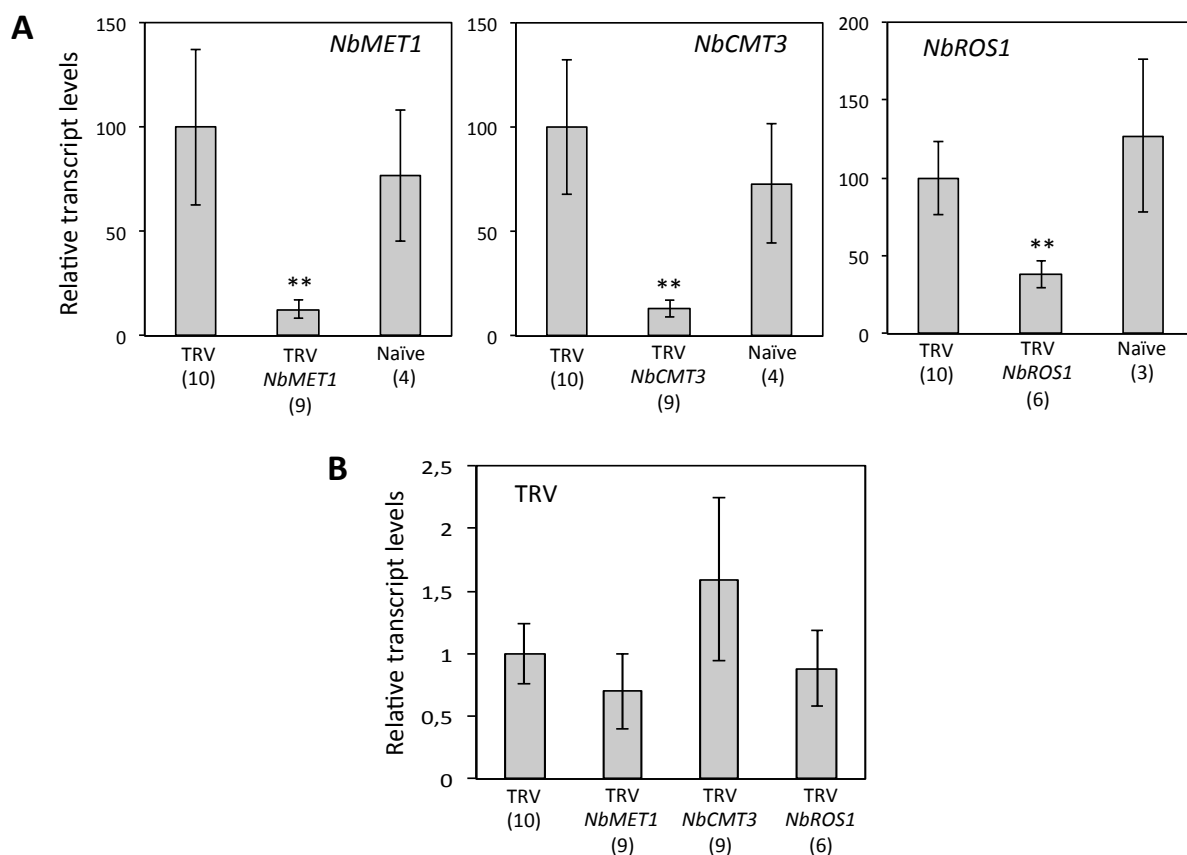
Figure 5. Relative transcript levels of *NbMET1, NbCMT3, NbROS1* and TRV determined by RT-qPCR after VIGS in *N. benthamiana* plants. Two fully expanded leaves of six-week-old *N. benthamiana* plants were agroinfiltrated with TRV-*NbMET1,* TRV-*NbCMT3*, TRV-*NbROS1* and empty TRV as control. At 18 dpi, total RNA was extracted from the pooled second and third most apical leaves and relative transcript levels of **(A)** *NbMET1, NbCMT3, NbROS1* and **(B)** TRV were determined by RT-qPCR. Expression levels were normalized to *NbEF1α* and are presented as the relative expression compared with TRV plants (set to 100 in **(A)** and set to 1 in **(B)**). Bars represent mean values ± standard deviation of two independent biological replicates, with a total of (n) plants. Asterisks indicate significant differences according to Student's t-test (**, p-value < 0.01).

We performed another biological replicate similar to that shown in Figure 5 to determine the *NbMET1, NbCMT3, NbDRM2* and *NbROS1* transcript levels not only in the corresponding silenced plants, but also in plants in which one of the other genes (either *NbMET1, NbCMT3, NbDRM2* or *NbROS1*) was silenced. *N. benthamiana* plants were agroinfiltrated with TRV (negative control), TRV-*NbMET1*, TRV-*NbCMT3*, TRV-*NbDRM2* and TRV-*NbROS1*, and their relative transcript levels were determined by RT-qPCR (Figure 6). For *NbMET1, NbCMT3* and *NbROS1*, a transcript level reduction of 75% (in TRV-*NbMET1*, bar is shown in black), 43% (in TRV-*NbCMT3*, bar is shown in black) and 46% (in TRV-*NbROS1*, bar is shown in black), respectively, was observed when compared with TRV-treated plants (Figure 6), while as described in previous experiments, the silencing of *NbDRM2* was not achieved (data not shown). We could not detect any significant changes in the expression of the other components of the DNA methylation machinery when either *NbMET1, NbCMT3* or *NbROS1* were silenced (Figure 6).

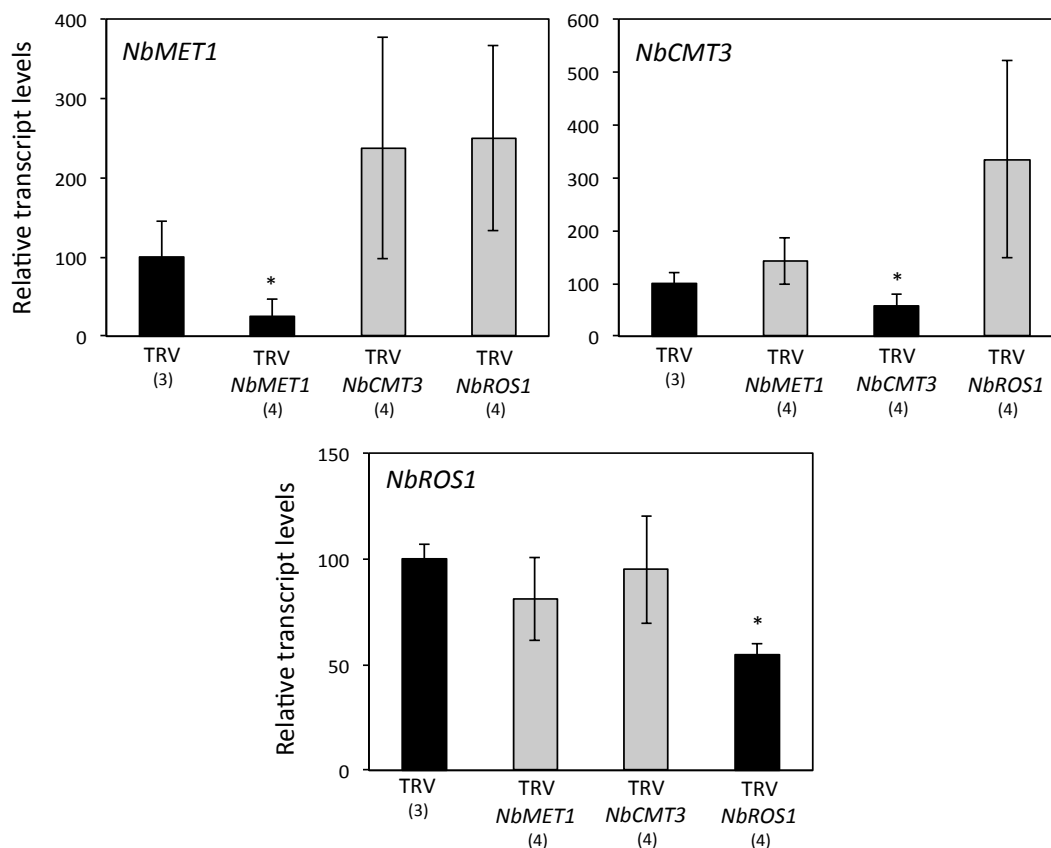**Figure 6. Relative transcript levels of *NbMET1, NbCMT3* and *NbROS1* determined by RT-qPCR after virus-induced gene silencing (VIGS) in *N. benthamiana* plants.** Two fully expanded leaves of six-week-old *N. benthamiana* plants were agroinfiltrated with TRV-*NbMET1,* TRV-*NbCMT3*, TRV-*NbROS1* and empty TRV as control. At 18 dpi, total RNA was extracted from the pooled second and third most apical leaves and relative transcript levels of *NbMET1, NbCMT3* and *NbROS1* were determined by RT-qPCR. Gene expression levels were normalized to *NbEF1α* and are presented as the relative expression compared with TRV plants (negative control) (set to 100). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Asterisks indicate significant differences according to Student's t-test (*, p-value < 0.05). Results presented in this Figure were obtained by Dr. Edgar Rodríguez Negrete.

Additionally, we evaluated if the down-regulation of the DNA methylation machinery by VIGS was intense enough to have an impact on the plant cytosine methylation landscape. Total DNA was extracted from the same tissue that was previously characterized by RT-qPCR (Figure 6), and was digested with two methylation-sensitive enzymes: *Msp*I (cuts when CG and/or CHG methylation is lost) and *Hpa*II (cuts when CG methylation is lost). Digested DNA was loaded on an agarose gel (Figure 7A) and transferred to a nylon membrane to perform a Southern blot (Figure 7B). As a radiolabeled probe, the *N. benthamiana HRS60* sequence, a mostly subtelomeric 183 pb tandem repeat (1.6×10^5 copies, Koukalová et al., 1989) that shows high levels of CG and CHG methylation (Kovarik et al., 2000), was used. The *Msp*I-digested DNA from TRV-*NbMET1*- and TRV-*NbCMT3*-inoculated plants yielded an increase in the accumulation of DNA corresponding to monomers, dimers and multimers of *NbHRS60*, which indicates the genomic DNA from these silenced plants was more readily digested in the silenced plants (Figure 7B). Similarly, the DNA from TRV-*NbMET1*-inoculated plants also showed an increase in the number of monomers and multimers when digested with *Hpa*II, indicating that hypomethylation at the CG context occurs

when *NbMET1* levels are diminished (Figure 7B). On the other hand, no differences were detected when comparing the restriction pattern of digested DNA from TRV control and TRV-*NbROS1*- or TRV-*NbDRM2*-inoculated plants (Figure 7B).
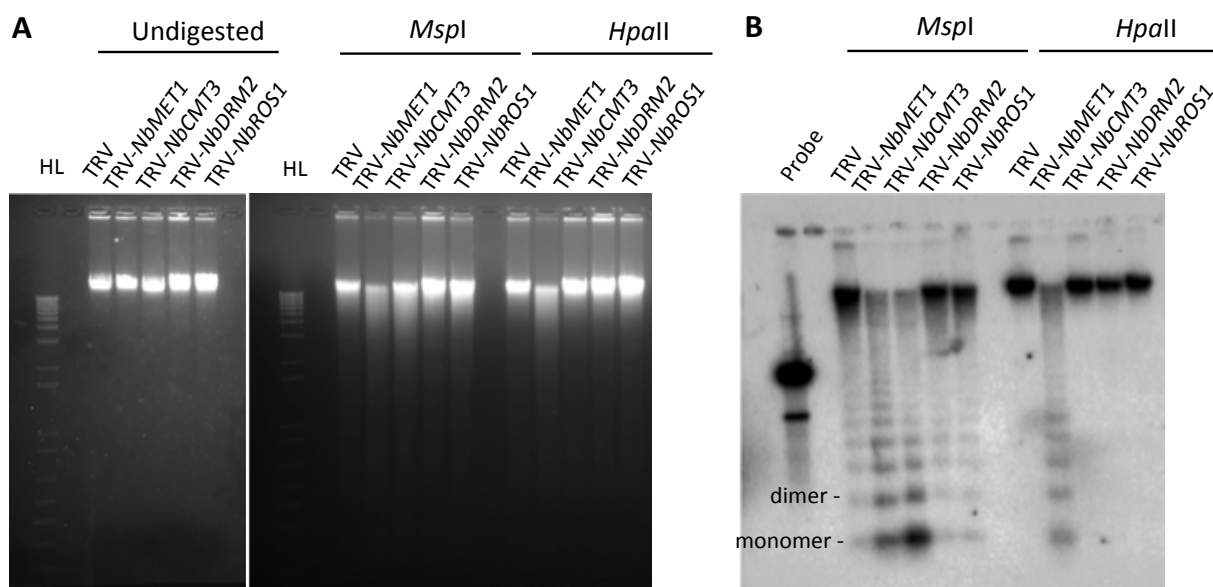

Figure 7. *NbHRS60* DNA accumulation in *N. benthamiana* plants inoculated with TRV-*NbMET1*, TRV-*NbCMT3*, TRV-*NbDRM2* or TRV-*NbROS1*, after digestion with methylation-sensitive enzymes. Total DNA was extracted from the same tissue used to test the silencing (Figure 6), and digested with *Msp*I (cuts when CG and/or CHG methylation are lost, where H is A, T or C) or *Hpa*II (cuts only when CG methylation is lost). **(A)** Ethidium bromide-stained digested and non-digested DNA. **(B)** Digested DNA was transferred to a nylon membrane and hybridized with a radiolabeled specific probe for *NbHRS60*. Bands correspond to *NbHRS60* monomers (183 bp), dimers and multimers. Each lane corresponds to one representative plant. Results presented in this Figure were obtained by Dr. Edgar Rodríguez Negrete.

We can conclude that the TRV-*NbMET1* and TRV-*NbCMT3* constructs significantly silenced the specific *N. benthamiana* genes (Figure 6), and this silencing induced hypomethylation at CG and CHG sites of the heterochromatic subtelomeric repeat *NbHRS60* (Figure 7B). Relative transcript levels of *NbROS1* were also significantly reduced when VIGS was performed using TRV-*NbROS1* (Figure 6) and, as expected, no differences were observed in the DNA methylation pattern (Figure 7B) as the homologue of this DNA demethylase in *A. thaliana* is involved in the active demethylation of just a few hundred of *loci* (Lister et al., 2008). Unfortunately, we were not able to silence *NbDMR2* with the designed construct and we decided not to include it in future experiments.

A previous work from our group used a reverse genetic approach to identify host factors that were necessary for geminivirus infection (Lozano-Durán et al., 2011). This work showed that inducing gene silencing by VIGS, agroinoculating the geminivirus at the same time point and sampling the most apical tissue around two to three weeks after the infection, delivered a good compromise between inducing the silencing of the host gene by TRV and letting the geminivirus replicate in this modified cell environment. In order to determine the biological relevance of NbMET1-, NbCMT3- and NbROS1-deficiency in TYLCV infection in *N. benthamiana*, we infected

six-week-old plants by agroinoculating TYLCV in the axillary bud of the fourth/fifth leaf. At the same time point, two fully expanded leaves were agroinfiltrated with TRV, TRV-*NbMET1*, TRV-*NbCMT3* or TRV-*NbROS1*. To determine the silencing degree of TRV-*NbMET1*, TRV-*NbCMT3* and TRV-*NbROS1*, total RNA was extracted at 18 dpi from the second and third most apical leaves (pooled), and the relative transcript accumulation of the silenced genes was determined using RT-qPCR (*NbEF1α* transcript levels were used as normalizer). Total DNA was extracted from the same tissue (second and third most apical leaves) and relative TYLCV DNA levels were measured by qPCR. DNA levels were normalized to the *25S* ribosomal gene, previously used as endogenous gene for geminiviral quantification (Mason et al., 2007; Lozano-Durán et al., 2011; Rodríguez-Negrete et al., 2014). TRV-*NbMET1*- and TRV-*NbROS1*-agroinfiltrated plants were analysed in an independent experiment from TRV-*NbCMT3*-inoculated plants.
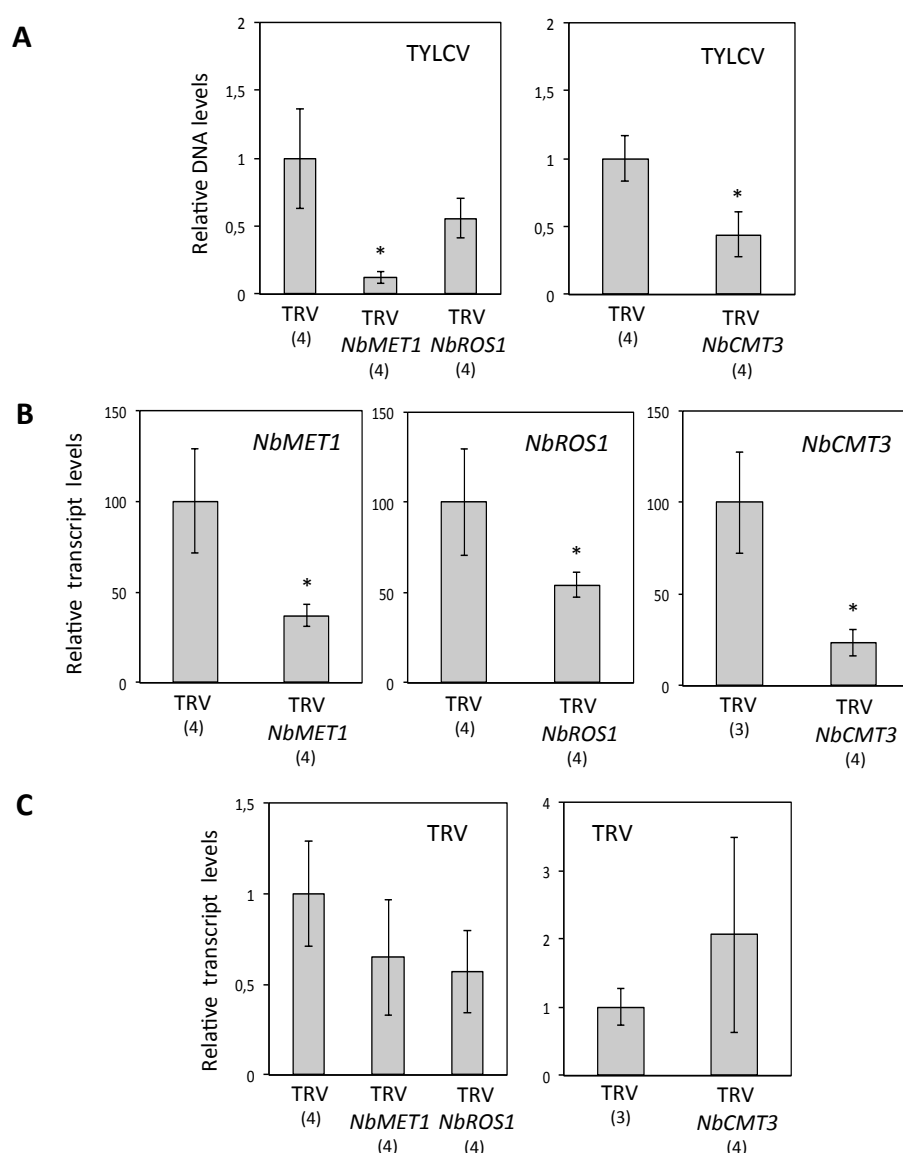

Figure 8. Relative TYLCV accumulation and relative transcript levels of *NbMET1*, *NbROS1* and *NbCMT3* determined by qPCR in *NbMET1*-, *NbROS1*- and *NbCMT3*-silenced *N. benthamiana* plants. Two fully expanded leaves of six-week-old *N. benthamiana* plants were agroinfiltrated with TRV-*NbMET1,* TRV-*NbROS1*, TRV-*NbCMT3* and empty TRV as control. The axillary bud was agroinoculated with the geminivirus TYLCV. **(A)** At 18 dpi, total DNA was extracted from the pooled second and third most apical leaves and relative TYLCV DNA levels were determined by qPCR. DNA

levels were normalized to *Nb25S* and are presented as the relative accumulation compared with TRV plants (set to 1). At 18 dpi, total RNA was extracted from the same tissue than DNA and the relative transcript levels of **(B)** *NbMET1, NbROS1, NbCMT3* and **(C)** TRV were determined by RT-qPCR. Expression levels were normalized to *NbEF1α* and are presented as the relative expression compared with TRV plants (set to 100 in **(B)** and set to 1 in **(C)**). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Asterisks indicate significant differences according to Student's t-test (*, p-value < 0.05).

We observed a relative transcript level reduction of 63%, 46% and 77% for *NbMET1*, *NbROS1* and *NbCMT3*, respectively, compared with non-silenced control plants (TRV) (Figure 8B), with no significant differences of TRV accumulation in all the plants assayed (Figure 8C). *NbMET1*-silenced plants displayed a geminiviral load reduction of 88%, suggesting an important role for NbMET1 during TYLCV infection (Figure 8A). On the other hand, TYLCV accumulation was reduced in *NbROS1*-silenced plants when compared with TRV plants, although this difference was not statistically significant (Figure 8A). In TRV-*NbCMT3*-agroinfiltrated plants, a significant TYLCV viral load reduction of 56% was detected (Figure 8A).

These results suggested that the down-regulation of the *N. benthamiana* maintenance DNA methylation machinery, *NbMET1* and *NbCMT3*, impairs TYLCV infection. This is not consistent with what we observed for TYLCV infection in *A. thaliana* mutants lacking MET1 or CMT3 (Figure 1 and 2), and with previous results from Sun et al. (2015), which described that the accumulation of the geminivirus *Indian cassava mosaic virus* (ICMV) is higher in *NbCMT3*-silenced *N. benthamiana* plants. Therefore, we further analysed TYLCV DNA levels (Figure 8) by checking carefully the qPCR-derived Ct values. The Ct (also named Cq) value is the cycle at which the fluorescence generated within a qPCR reaction crosses the fluorescence threshold. This cycle is inversely proportional to the original DNA amount of the gene of interest.

A

| Gene (DNA) | Ct value for DNA amount | | | | | | | | | | | |
| | TRV | | | | TRV-*NbMET1* | | | | TRV-*NbROS1* | | | |
	P1	P2	P3	P4	P1	P2	P3	P4	P1	P2	P3	P4
TYLCV	12.99	13.25	13.48	13.15	12.64	13.12	13.33	13.81	12.75	13.73	12.85	13.13
Nb25S	21.38	21.20	21.30	20.18	18.06	17.93	17.53	18.46	20.24	20.29	19.91	20.15
NbACT	24.64	24.06	24.97	24.60	24.78	24.70	24.89	25.50	25.00	25.24	24.85	24.98
NbEF1α	24.52	24.13	24.95	24.63	24.33	24.24	24.68	25.21	24.51	24.87	24.34	24.63

B

| Gene (DNA) | Ct value for DNA amount | | | | | | | |
| | TRV | | | | TRV-*NbCMT3* | | | |
	P1	P2	P3	P4	P1	P2	P3	P4
TYLCV	11.45	12.30	11.88	11.96	11.21	11.15	12.02	12.43
Nb25S	19.67	20.19	20.11	19.64	17.90	18.57	18.84	18.61
NbACT	23.62	23.57	23.63	23.66	23.06	23.26	23.60	23.73
NbEF1α	22.75	22.59	22.99	23.07	22.60	22.40	23.01	23.21

C

| Gene (DNA) | Ct value for DNA amount | | |
| | TRV | TRV-*NbMET1* | TRV-*NbROS1* |
	Mean ± SD	Mean ± SD	Mean ± SD
TYLCV	13.22 ± 0.21	13.22 ± 0.49	13.11 ± 0.44
Nb25S	21.01 ± 0.56	17.99 ± 0.38	20.14 ± 0.17
NbACT	24.58 ± 0.38	24.96 ± 0.36	25.01 ± 0.17
NbEF1α	24.55 ± 0.34	24.61 ± 0.44	24.59 ± 0.22

D

| Gene (DNA) | Ct value for DNA amount | |
| | TRV | TRV-*NbCMT3* |
	Mean ± SD	Mean ± SD
TYLCV	11.90 ± 0.35	11.7 ± 0.63
Nb25S	19.91 ± 0.29	18.48 ± 0.40
NbACT	23.62 ± 0.04	23.41 ± 0.30
NbEF1α	22.85 ± 0.21	22.81 ± 0.37

Table 1. Ct values for TYLCV, *Nb25S*, *NbACT* and *NbE1Fα* DNA in *NbMET1*-, *NbROS1*- and *NbCMT3*-silenced *N. benthamiana* plants. (A, B) Ct values corresponding to TYLCV, *Nb25S, NbACT* and *NbEF1α* for each TRV-, TRV-*NbMET1*-, TRV-*NbROS1* and TRV-*NbCMT3*-inoculated plant (P). **(C, D)** Mean and SD of Ct values shown in **(A, B)**.

Table 1 depicts the Ct values or cycles corresponding to *Nb25S* (normalizer gene) and the TYLCV DNA, obtained by qPCR using the DNA from TRV-inoculated plants (Figure 8). Table 1A and 1B shows the cycle for each plant assayed (P), and Table 1C and 1D, the mean and standard deviation (SD). Strikingly, the Ct values of TYLCV were similar when comparing non-silenced (TRV) and silenced plants (TRV-*NbMET1*, TRV-*NbCMT3* and TRV-*NbROS1*), but there was a decrease of more than 1.4 cycles for *Nb25S* in TRV-*NbMET1*- and TRV-*NbCMT3*-silenced plants compared with non-silenced ones (TRV, Table 1). This suggests that the silencing of the maintenance DNA methyltransferases could alter the amount of *Nb25S* DNA that is detected by qPCR. As we have used *Nb25S* to relativize the amount of TYLCV found in the TRV-silenced plants, this observation could be misleading our results. Therefore, we determined the Ct values for two other plant *loci, Actin (NbACT)* and *NbEF1α*, in those same samples, and found no significant differences in their Ct values between non-silenced plants (TRV) and plants in which the DNA methylation machinery had been silenced (Table 1). When we measured the relative amount of TYLCV DNA using any of these two normalizers, *NbACT* or *NbEF1α*, we could no longer detect the reduction of TYLCV levels when the maintenance DNA methyltransferases, *NbMET1* or *NbCMT3*, are silenced (Figure 9).
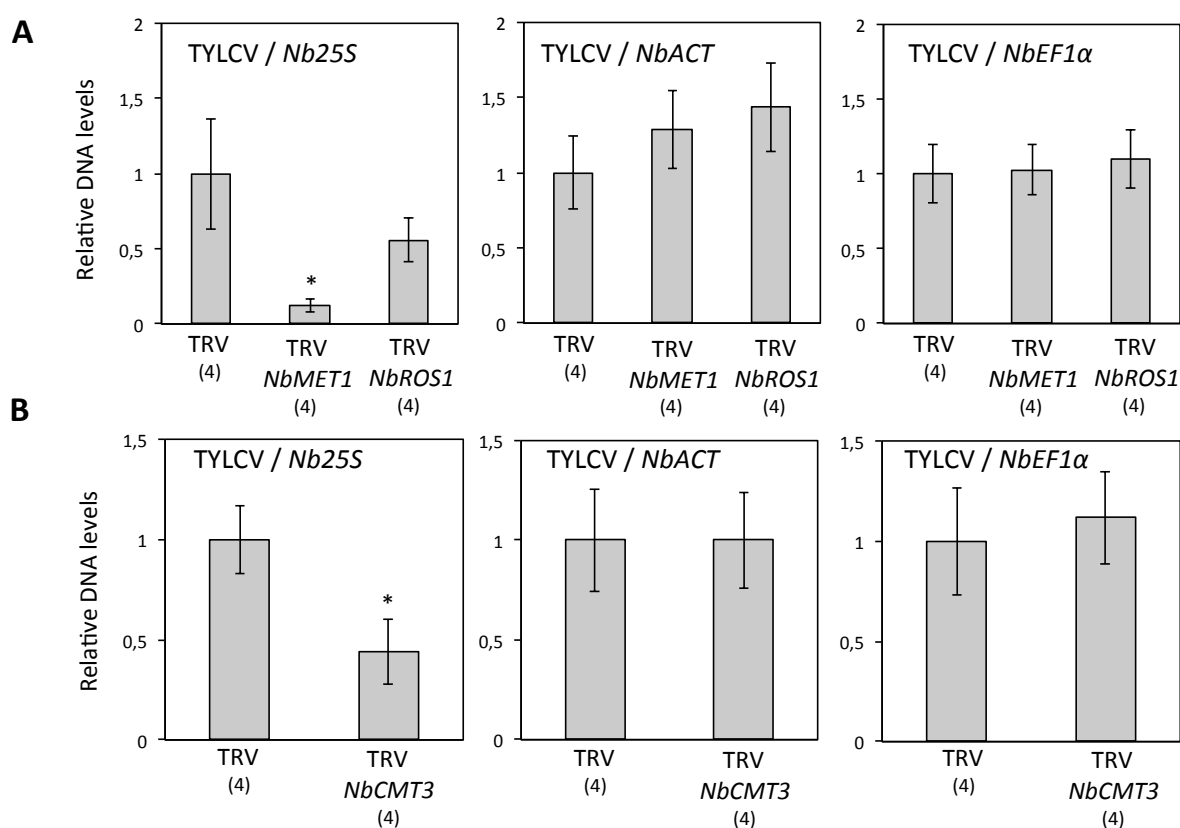

Figure 9. Relative TYLCV accumulation determined by qPCR in *NbMET1*-, *NbROS1*- and *NbCMT3*-silenced *N. benthamiana* plants. Relative TYLCV levels determined by qPCR in **(A)** TRV-*NbMET1*, TRV-*NbROS1*, **(B)** TRV-*NbCMT3* and **(A, B)** TRV plants, using *Nb25S, NbACT* or *NbEF1α* as normalizer (TYLCV/*Nb25S*, TYLCV/*NbACT* or TYLCV/*NbEF1α*). DNA levels are presented as the relative accumulation compared with TRV plants (set to 1). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Asterisks indicate significant differences according to Student's t-test (*, p-value < 0.05).

In order to confirm if the reduction of TYLCV DNA levels, detected in plants impaired in the maintenance DNA methyltransferases, was an artefact due to the effect that the *NbMET1* or *NbCMT3* silencing may have on the detection of *Nb25S* DNA by qPCR, two biological replicates following the same procedure shown in experiments from Figure 8 were performed. *N. benthamiana* plants were agroinfiltrated at the same time with TYLCV and the TRV vectors to silence either *NbMET1* or *NbCMT3*, and total RNA and DNA were extracted from the same tissue samples at 18 dpi. The relative transcript levels of the silenced genes and TRV, and the level of TYLCV accumulation, were measured by RT-qPCR and qPCR, respectively. *NbEF1α* was used as a normalizer for RT-qPCR, and *NbACT* to normalize TYLCV DNA levels. Just one representative biological replicate is shown in Figure 10, which confirmed our previous finding: the down-regulation of *NbMET1* or *NbCMT3* does not have an effect on TYLCV accumulation in *N. benthamiana* plants.
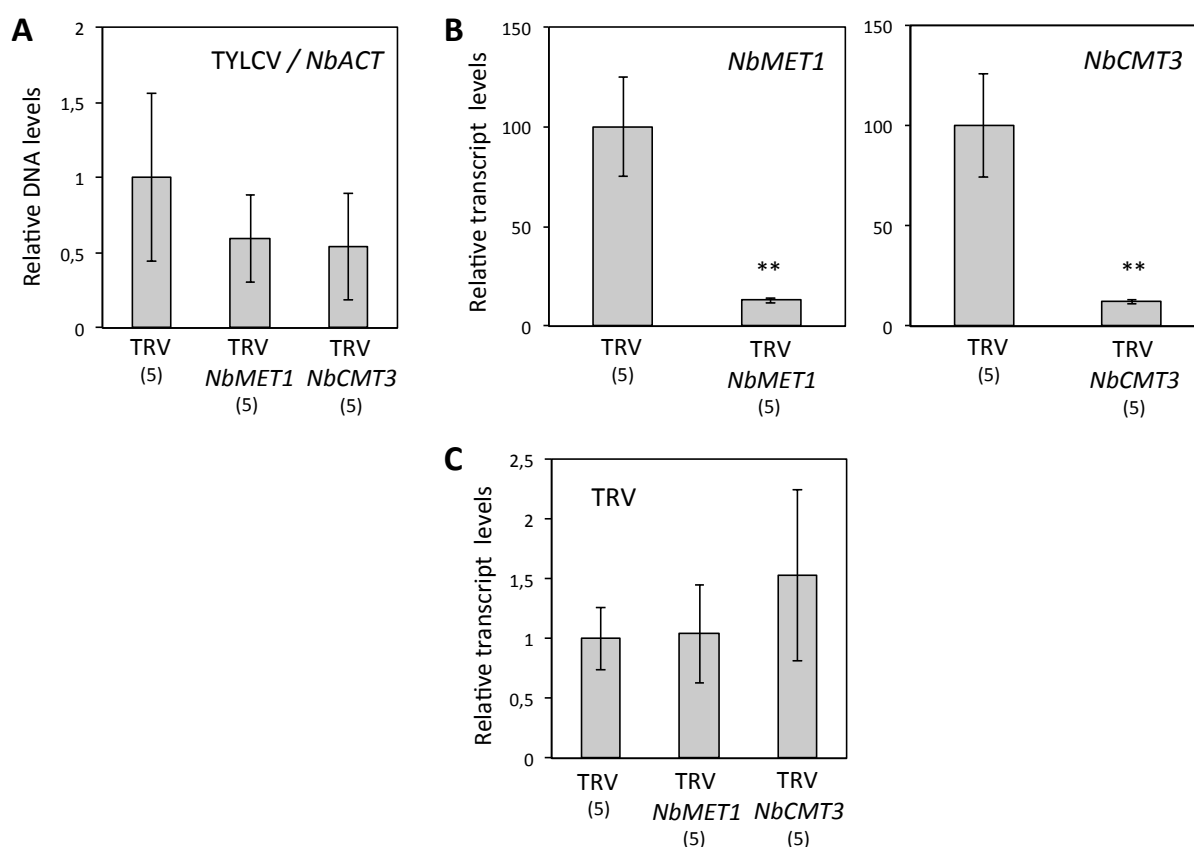
Figure 10. Relative TYLCV accumulation and relative transcript levels of *NbMET1* and *NbCMT3* determined by qPCR in *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants. Two fully expanded leaves of six-week-old *N. benthamiana* plants were agroinfiltrated with TRV-*NbMET1,* TRV-*NbCMT3* and empty TRV as control. The axillary bud was agroinoculated with the geminivirus TYLCV. **(A)** At 18 dpi, total DNA was extracted from the second and third most apical leaves (pooled) and relative TYLCV DNA levels were determined by qPCR. DNA levels were normalized to *NbACT* and are presented as the relative accumulation compared with TRV plants (set to 1). At 18 dpi, total RNA was extracted from the same tissue than DNA and the relative transcript levels of **(B)** *NbMET1, NbCMT3* and **(C)** TRV were determined by RT-qPCR. Expression levels were normalized to *NbEF1α* and are presented as the relative expression compared with TRV plants (set to 100 in **(B)** and set to 1 in **(C)**). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Asterisks indicate significant differences according to Student's t-test (**, p-value < 0.01).

Based on these results, we conclude that *Nb25S* is not a valid normalizer for DNA measurements when the DNA methylation machinery is being silenced, as the impairment of DNA methylation alters the detection of *Nb25S* DNA. Therefore, according to data obtained using *NbACT* or *NbEF1α* as normalizers in three biological replicates, we can conclude that the down-regulation of *NbMET1*, *NbCMT3* and *NbROS1* did not impact TYLCV accumulation in *N. benthamiana.* Interestingly, the DNA amount of *Nb25S* seemed to be increased somehow, upon *NbMET1* or *NbCMT3* repression, and this preliminary result was further investigated in Chapter 3.

Finally, we assessed the role of NbMET1 and NbCMT3 on the infection with a second begomovirus, *Tomato yellow leaf curl* Sardinia *virus* (TYLCSV). We infected 9 to 13 *N. benthamiana* plants that were inoculated with TRV, TRV-*NbMET1* or TRV-*NbCMT3* in three independent biological replicates, following the same procedure than previously depicted for TYLCV infection assays. At 18 dpi, total RNA and DNA were extracted from the same tissue samples, and relative transcript levels of *NbMET1*, *NbCMT3* and TRV, and the level of TYLCSV DNA accumulation, were measured by RT-qPCR or qPCR, respectively. *NbEF1α* and *NbACT* were used to normalize the transcript and DNA levels, respectively.
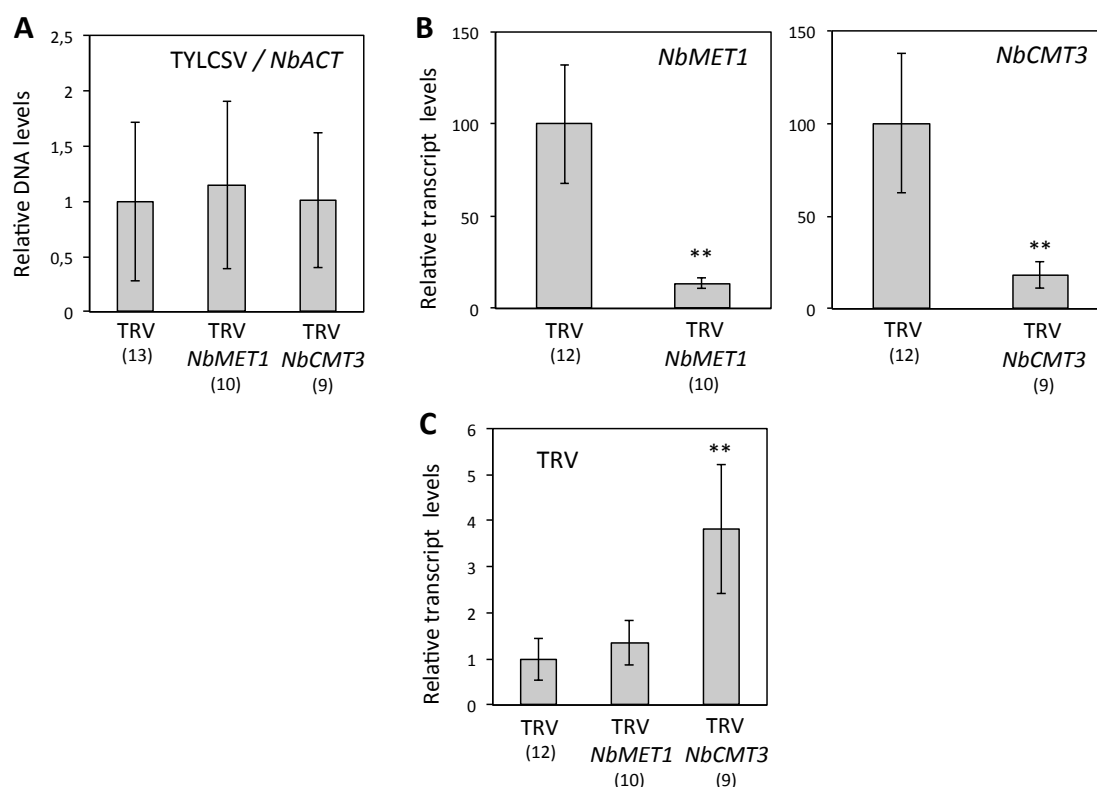

Figure 11. Relative TYLCSV accumulation and relative transcript levels of *NbMET1* and *NbCMT3* determined by qPCR in *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants. Two fully expanded leaves of six-week-old *N. benthamiana* plants were agroinfiltrated with TRV-*NbMET1,* TRV-*NbCMT3* and empty TRV as control. The axillary bud was agroinoculated with the geminivirus TYLCSV. **(A)** At 18 dpi, total DNA was extracted from the second and third most apical leaves (pooled) and relative TYLCSV levels were determined by qPCR. DNA levels were normalized to *NbACT* and are presented as the relative accumulation compared with TRV plants (set to 1). At 18 dpi, total RNA was extracted from the same tissue than DNA and the relative transcript levels of **(B)** *NbMET1, NbCMT3* and **(C)** TRV were determined by RT-qPCR. Expression levels were normalized to *NbEF1α* and are presented as the relative expression compared with TRV plants (set to 100 in **(B)** and set to 1 in **(C)**). Bars represent mean values ± standard deviation of three independent biological replicates, with a total (n) plants. Asterisks indicate significant differences according to Student's t-test (**, p-value < 0.01).

The mean values of the three biological replicates are shown in Figure 11, which let us conclude that, as observed in the previous TYLCV infection assays, the down-regulation of *NbMET1* or *NbCMT3* does not impact TYLCSV accumulation in *N. benthamiana*.

DISCUSSION

Previous results from our group demonstrated that *MET1* and *CMT3* transcript levels were down-regulated during geminiviral infection in *N. benthamiana*, and that the viral protein Rep was necessary for the repression of these maintenance DNA methyltransferases in *N. benthamiana* and *A. thaliana* (Rodríguez-Negrete *et al.*, 2013, Appendix II from this thesis). In this Chapter, we have evaluated the biological importance of the host DNA methylation machinery on geminivirus infection. The results obtained with the begomovirus TYLCV indicate that the lack of ROS1 (the main demethylase in vegetative tissues), MET1 (responsible for the maintenance of CG methylation), or CMT3 and DRM2 (responsible for the maintenance of CHG methylation and the *de novo* methylation in the three cytosines contexts, respectively) did not impact TYLCV accumulation in *A. thaliana*. Conversely, the work from Raja and colleagues (2008) showed that the *A. thaliana met1-7* and *cmt3-7* single mutants displayed enhanced disease symptoms for two other geminiviruses, BCTV and CaLCuV, which suggests an antiviral activity for the host maintenance DNA methylation machinery. Strikingly, they checked the geminivirus accumulation by Southern blot in floral tissue and found no differences in viral amount between wild-type and *met1-7* and *cmt3-7* mutant plants (Raja et al., 2008). This observation is in agreement with our results for TYLCV and it will argue against the role of DNA methylation as a defence mechanism against geminivirus. In order to bring some light into this controversy, we decided to quantify the accumulation levels of TYLCV and BCTV (the curtovirus used by Raja and colleagues, 2008) in the triple mutant background, *ddc*, which lacks CMT3, DRM1 and DRM2, and therefore is impaired in maintenance of DNA methylation at CHG context and in *de novo* DNA methylation.

As the increase of symptoms shown in the DNA methylation mutants after BCTV infection was mainly reported in the inflorescences (Raja et al., 2008) and we found that BCTV replicated and accumulated to a greater extent in floral tissue than in the rosette leaves, we determined the relative amount of virus in the *ddc* mutant in both tissues. We could not detect a clear increase of TYLCV levels in *ddc* mutant plants, probably due to the great variability in DNA levels and difficulty to infect that this asymptomatic virus shows in *A. thaliana*. On the other hand, we detected higher BCTV levels in the floral and leaf tissues of *ddc* mutant compared to Col-0 plants, although just the differences found in the inflorescences were statistically significant. This result was in agreement with previously published data and suggested that indeed the plant DNA methylation machinery interferes with curtovirus/geminivirus accumulation and may be playing an antiviral defence role. Repeating these experiments using another begomovirus that infects *A.*

thaliana more successfully than TYLCV, such as CaLCuV, will probably deliver more convincing results about the importance of the DNA methylation machinery in begomovirus infection.

We decided to study the importance of the plant DNA methylation machinery on TYLCV infection in *N. benthamiana*, as TYLCV induces symptoms, replicates and accumulates to a great extent in this host. A previous work from our group used a reverse genetic approach to identify host factors in *N. benthamiana* that were necessary for geminivirus infection (Lozano-Durán et al., 2011). Therefore, we silenced the genes encoding the DNA methyltransferases NbMET1, NbCMT3 and NbDRM2, and the DNA demethylase NbROS1, using VIGS, and checked the viral levels once the silencing was established. We confirmed that the expression was down-regulated to at least 40% for all of the genes, except for *NbDRM2*, which was discarded for further analysis. Moreover, we demonstrated that the silencing of *NbMET1* or *NbCMT3* by VIGS induced the hypomethylation of a heterocromatic sub-telomeric repeat such as *NbHRS60*. We observed similar TYLCV levels in the apical leaves from non-silenced control and *NbROS1*-, *NbMET1*- or *NbCMT3*-silenced plants, which suggested that the DNA methylation machinery, in *N. benthamiana*, does not play a role in TYLCV infection. The same result was obtained for the related specie TYLCSV. Conversely, the work from Sun et al. (2015) described that TRV-*NbCMT3*-silenced *N. benthamiana* plants displayed a greater accumulation of the bipartite begomovirus ICMV. The fact that they infected the plants with ICMV 10 days after inducing *NbCMT3* silencing and not at the same time point, could explain the differences found between both results. So these results seem to be contradictory regarding the biological relevance of the DNA methylation machinery in geminivirus infection, which may depend on the virus specie or the host used to determine such relevance. Moreover, it is worth noticing that other scientists from the geminivirus community argue that these viruses could evade repressive DNA methylation and transcriptionally gene silencing, simply via efficient Rep-dependent replication. Therefore, the biological relevance on geminivirus infection of the plant transcriptional gene silencing machinery will be very limited (reviewed in Pooggin et al., 2013).

Unexpectedly, we detected an increase in the amount of DNA for the ribosomal gene, *Nb25S*, when *NbMET1* or *NbCMT3* were silenced. This unexpected effect on the *Nb25S* DNA levels initially misled our interpretation of the impact that the silencing of these genes had on TYLCV accumulation. *Nb25S* was first used as a qPCR DNA normalizer for the quantification of geminivirus during the infection by Mason and colleagues (2007). Since then, other groups have used it to normalize the viral levels in infected samples (Rodríguez-Negrete et al., 2014) even when other host genes were silenced by VIGS (Lozano-Durán et al., 2011) and found no variation on the DNA amount for the *Nb25S* gene. As the plant genome harbours hundreds to thousands of copies of the rDNA genes that are tandemly repeated at heterochromatic regions (subtelomeric regions for the *45S* rDNA unit and pericentromeric for the *5S* rDNA unit), we have deeply characterized the effect that the silencing of the DNA methylation machinery could have on the stability of the plant rDNA genes on Chapter 3.

Supplementary data.

Genotyping of met1-3 plants.

The *met1-3* mutant plants were selected on medium containing 15 µg/ml phosphinotricin (Saze et al., 2003). Total DNA was extracted from 20 plants and analysed by PCR using the primers described in Johnson et al. (2007) (Figure S1A, Table 2 in Materials and methods). As a control for primers specificity, total DNA from Col-0 plants was amplified with the same primers. All analysed mutant plants were heterozygous for *met1-3* mutation, and a unique band corresponding to the wild-type allele was observed in Col-0 (Figure S1B). The heterozygous *met1-3* mutant was previously described to induce DNA hypomethylation, to reactivate transcriptionally silent *loci* and to induce epigenetically-controlled phenotypes, such as late flowering (Saze et al., 2003), suggesting that the lack of one of the *MET1* alleles is enough to revert transcriptional gene silencing.
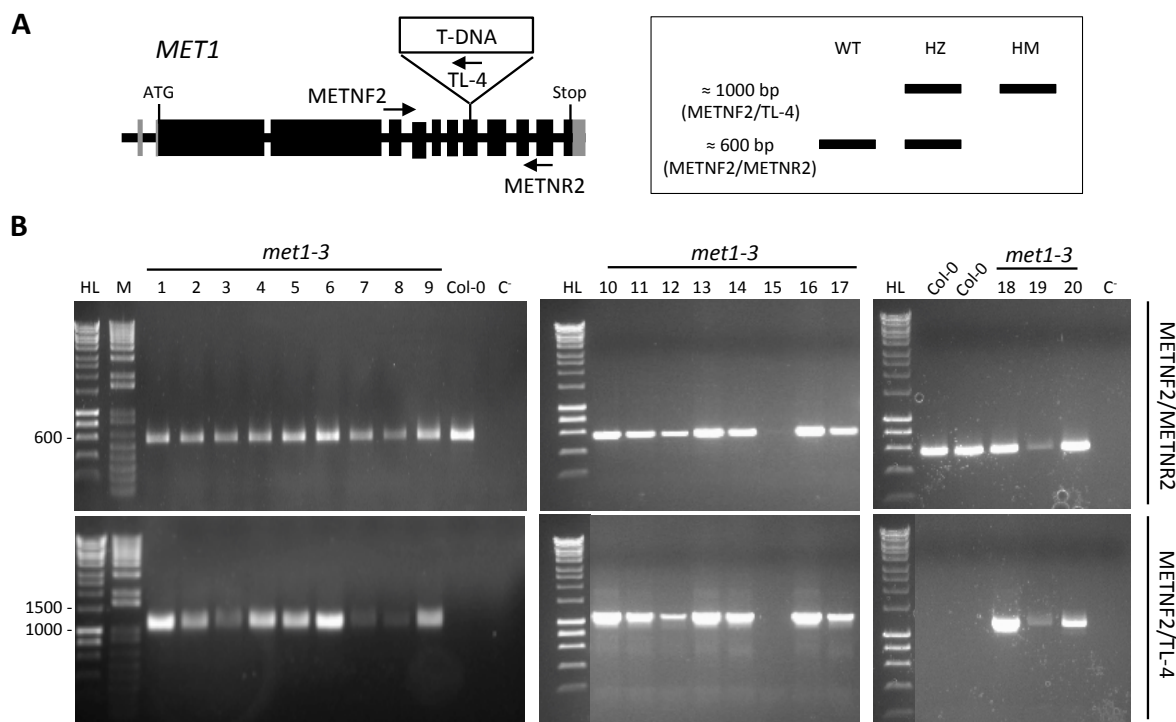
Figure S1. Genotyping *A. thaliana* met1-3 mutant. (A) Primers used to amplify the wild-type (METMF2/METNR2 primers pair) and the mutant allele (METNF2/TL-4 primers pair) of *MET1*, and expected band sizes in wild-type Col-0 (WT), heterozygous mutant (HZ) and homozygous mutant (HM) plants (Johnson et al., 2007). Black boxes represent exons. **(B)** Agarose gels containing the PCR products using the primers depicted in **(A)** to amplify DNA from Col-0 and *met1-3* plants (samples 1 to 20). C-: non-template control. HL: HyperLadder, molecular weight marker. M: molecular weight marker.

Genotyping of ros1-4 plants.

Total DNA from five *ros1-4* plants was analysed by PCR using the primers described in SIGnAL (Salk Institute Genomic Analysis Laboratory, www.Signal.salk.edu/tdnaprimers.2.html) (Figure S2A, Table 2 in Materials and methods). A unique band corresponding to the wild-type allele was observed in the Col-0 plant and not in the five *ros1-4* plants assayed, which showed amplification for the T-DNA insertion at *ROS1* gene (Figure S2B). As the five plants from the same replica showed the same genotype, we concluded that the *ros1-4* line was homozygous for the mutation.
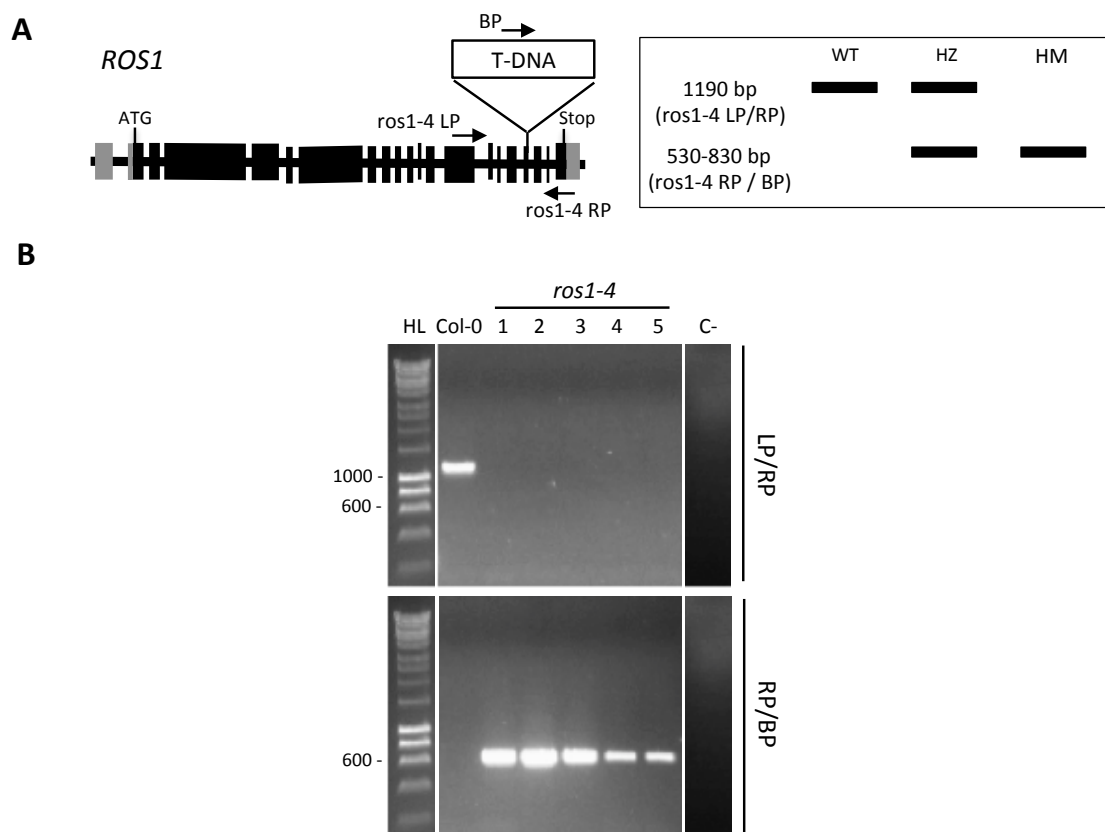
Figure S2. Genotyping *A. thaliana ros1-4* mutant. (A) Primers used to amplify the wild-type (ros1-4 LP/RP primers pair) and the mutant (ros1-4 RP / BP primers pair) allele of *ROS1*, and expected band sizes in wild-type Col-0 (WT), heterozygous mutant (HZ) and homozygous mutant (HM) plants (SIGnAL). Black boxes represent exons. **(B)** Agarose gel containing PCR products using primers depicted in **(A)** to amplify DNA from Col-0 and *ros1-4* plants (samples from 1 to 5). C-: non-template control. HL: HyperLadder, molecular weight marker. Results from plants used in one biological replicate are shown.

Genotyping of ddc plants.

Total DNA from four *ddc* plants was analysed using primers described in SIGnAL for the T-DNA insertions (Figure S3A, Table 2 in Materials and methods). All the *ddc* plants analysed were homozygous for *drm2-2* and *cmt3-11* mutations as they showed amplification for the T-DNA insertion at the corresponding gene and not for the wild-type allele (Figure S3B). Strikingly, these plants were wild-type for *DRM1* due to the amplification was just detected with primers for the

DRM1 wild-type allele (Figure S3B). We detected amplification of the wild-type allele of *DRM1*, *DRM2* and *CMT3* in Col-0 (Figure S3C). Therefore, we can conclude that the *ddc* seeds that were supplied by the NASC (Nottingham Arabidopsis Stock Centre) correspond to a homozygous double mutant *drm2-2 cmt3-11* (named *dc* and used in infection assays showed in Figure 2) and not the triple mutant *drm1-2 drm2-2 cmt3-11* (*ddc*).
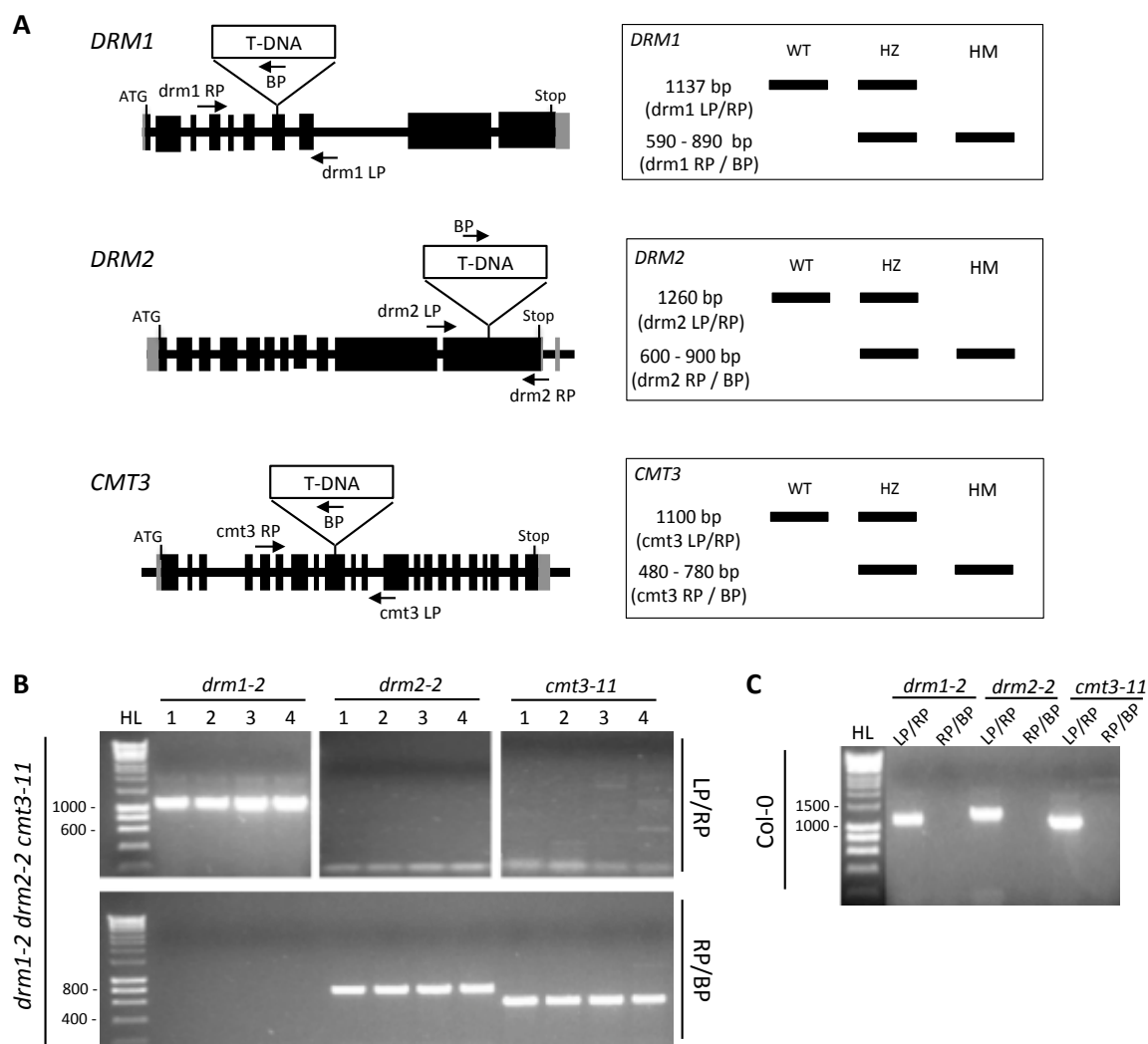

Figure S3. Genotyping *A. thaliana ddc* mutant. (A) Primers used to amplify the wild-type and the mutant allele, and expected band sizes in wild-type (WT), heterozygous mutant (HZ) and homozygous mutant (HM) plants (SIGnAL). Black boxes represent exons. **(B)** PCR products using primers depicted in **(A)** to amplify the wild-type allele of *DRM1* (drm1 LP/RP primers pair), *DRM2* (drm2 LP/RP) and *CMT3* (cmt3 LP/RP) from *ddc* plants (from 1 to 4), and to amplify the mutant allele of *DRM1* (drm1 RP / BP primer pair), *DRM2* (drm2 RP / BP) and *CMT3* (cmt3 RP / BP) from the same *ddc* plants. **(C)** DNA from one Col-0 plant amplified with the complete array of primers. C-: non-template control. HL: HyperLadder, molecular weight marker. Results from plants used in one biological replicate are shown.

Genotyping of the new set of ddc plants.

Col-0 and *ddc* plants were grown on soil and total DNA was extracted from rosette leaves, and the plants were genotyped by PCR with the primers previously indicated and shown in Figure S3A and Table 2 in Materials and methods.
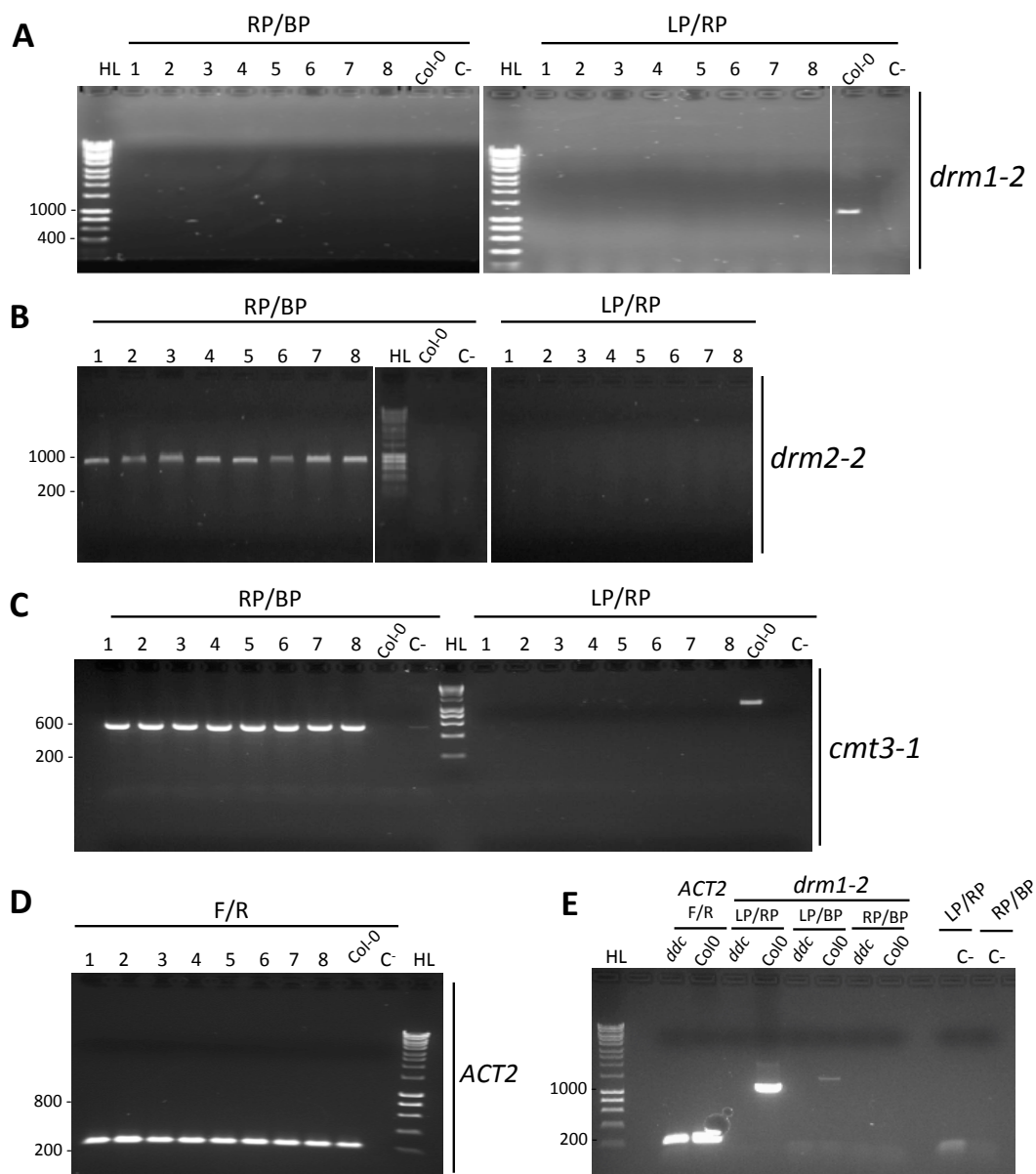

Figure S4. Genotyping *A. thaliana ddc* mutant. PCR products using primers depicted in Figure S3A to amplify the wild-type and the mutant allele of **(A)** *DRM1*, **(B)** *DRM2* and **(C)** *CMT3* in Col-0 and *ddc* (samples from 1 to 8) plants. **(E)** The BP and drm1 LP primers were used to amplify a possible T-DNA inversion. **(D)** ACT2 F/R primers were used as a control of template quality and quantity. C-: non-template control. HL: HyperLadder, molecular weight marker.

The PCR-based genotyping of the *drm2-2* and *cmt3-11* alleles showed that the new *ddc* plants were homozygous for both mutations (Figure S4B and S4C). However, no amplification was obtained for the wild-type or the mutant allele of *DRM1*, using the LP/RP and RP/BP primers

pairs, respectively (Figure S4A). We assessed that the quality and quantity of the DNA from Col-0 and *ddc* plants was adequate for PCR-amplification by using primers for *Actin* (*ACT2*) (Table 2 in Materials and methods), which yielded a 211 bp band in all the samples (Figure S4D). To assess if the T-DNA was inserted at the *DRM1* locus at the inverted orientation, we used the LP/BP primers, which did not amplify (Figure S4E).

The lack of amplification of both the wild-type and the mutant alleles in *ddc* plants along with the amplification of the *ACT2* gene, suggested that *DRM1* was mutated in these *ddc* plants, but the T-DNA might be located at a different *locus* from that predicted by SIGnAL. To address this question, we designed an array of primers pairs upstream and downstream the theoretical T-DNA insertion at the *DRM1* gene (Figure S5A), and total DNA from two Col-0 and two *ddc* plants was analysed by PCR.
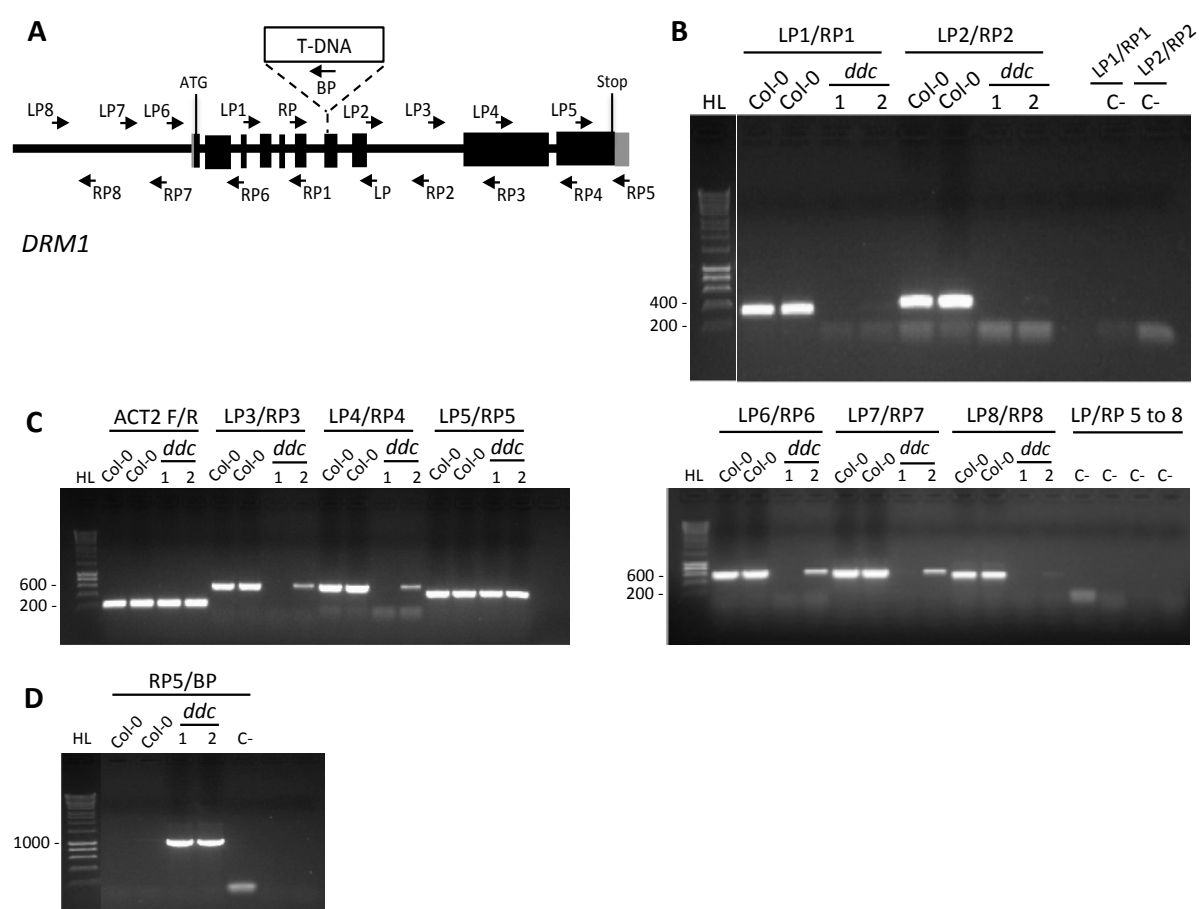

Figure S5. Genotyping *A. thaliana drm1* mutation. (A) Diagram of the *DRM1* ORF and an array of primers annealing on it. The theoretical position of the T-DNA is depicted by the dashed line. The number of the LP/RP primers pairs highlights how they have been paired for the amplification reaction. Black boxes represent exons. Grey box represents the 3'UTR. **(B, C)** PCR products using primers depicted in **(A)** to amplify the wild-type allele of *DRM1* in Col-0 and *ddc* (samples 1 and 2) plants. **(D)** PCR products using the BP and drm1 RP5 primers pair to amplify the mutant allele of *DRM1* in Col-0 and *ddc* plants. ACT2 F/R primers were used as a control of template quality and quantity. C-: non-template control. HL: HyperLadder, molecular weight marker.

All the primers pairs (LP/RP 1 to 4 and 6 to 8) yielded a band of the expected size in Col-0 but not in *ddc* plants, indicating that *DRM1* sequences were missing in the *ddc* mutant (Figure S5B and S5C). Only the primer pair LP5/RP5 showed amplification in *ddc* plants (Figure S5C), suggesting that the T-DNA insertion is just upstream the LP5 sequence. To check that the T-DNA insertion was upstream the LP5/RP5 region, we performed a PCR with the BP primer and the RP5 one, and we obtained a single band of approximately 1000 bp in *ddc* plants (Figure S5D), indicating that there was a T-DNA insertion at the *DRM1* locus. These results confirmed that *ddc* plants contain a T-DNA insertion at the *DRM1 locus* and that they are homozygous for the *drm1* mutation.

Chapter 3

Characterization of the ribosomal DNA (rDNA) in DNA methylation-deficient plants

BACKGROUND

The eukaryotic ribosome consists of two subunits formed by the intricate association of more than seventy ribosomal proteins and four distinct ribosomal RNAs (rRNAs). The small subunit (40S) comprises the *18S* rRNA, whereas the large subunit (60S) contains the *5S, 5.8S*, and *25S/28S* rRNAs (Yusupova et al., 2014). A single copy of the ribosomal DNA (rDNA) could not be transcribed at a sufficient rate for the cell's ribosomes requirement and thus, all organisms carry multiple rDNA copies, which are generally arranged in one or more tandem repeat arrays. However, the number of rDNA repeats is very variable throughout phylogeny and the reason that some organisms have thousands of repeats is yet unknown. The three rDNA genes, *18S, 5.8S* and *25S/28S,* are codified by the head-to-tail tandem repeats of the *45S* rDNA unit (~150 copies per haploid genome in yeast *Saccharomyces cerevisiae,* 300–400 in a diploid human cell and thousands of copies in higher plants). In yeast, the same rDNA repeat also encodes the *5S* rRNA, which is transcribed in the reverse direction, but in higher eukaryotes, the *5S* ribosomal RNA is located in different tandem arrays (reviewed in Shaw and McKeown, 2011). In *A. thaliana*, the *45S* rDNA clusters are positioned sub-terminally, at the distal ends of the short arms of chromosomes 2 and 4 (Kaul et al., 2000), while the approximately 1000 copies of *5S* rRNA genes per haploid genome are localized in the pericentromeric heterochromatin of chromosomes 3, 4 and 5 (Campell et al., 1992; Fransz et al., 1998; Murata et al., 1997). The genic regions of the rDNA units (*5S, 18S, 25S* and *5.8S* subunits) are highly conserved between species with respect to length and nucleotide sequence, but the intergenic spacer (IGS) between the rRNA gene repeats, rapidly diverges and can show inter- and intraspecific polymorphisms (Kovarik et al., 2008). Figure 1 shows the genomic structure of the *45S* and *5S* rDNA units from *N. benthamiana,* which are similar to the *A. thaliana* ones.
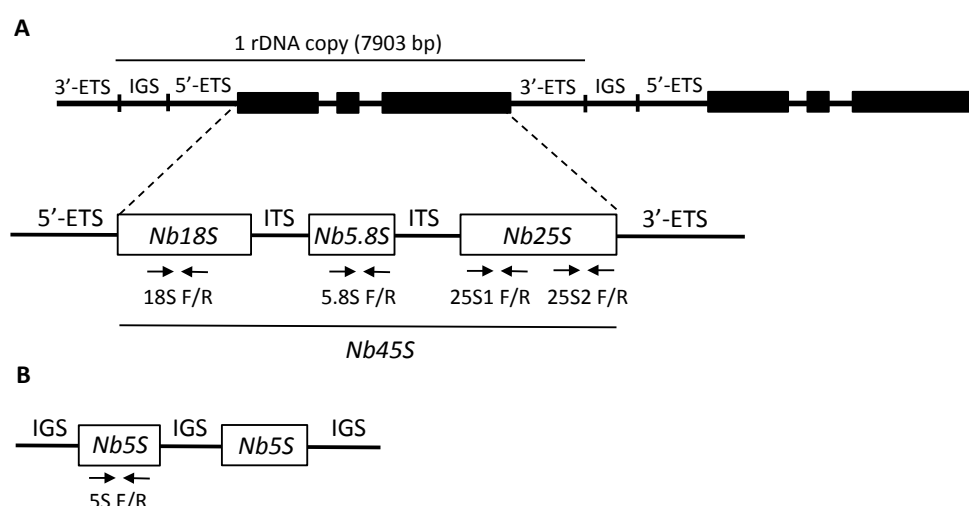

Figure 1. Diagram (not to scale) of the *N. benthamiana Nb45S* and *Nb5S* rDNA. (A) *Nb45S* rDNA is a subtelomeric region arranged in hundreds of tandemly repeated copies separated by non-transcribed intergenic spacers (IGS). Each copy (7903 bp) contains the *Nb18S, Nb5.8S* and *Nb25S* ribosomal genes separated by internal transcribed spacers (ITS), and two external transcribed spacers (5'-ETS and 3'-ETS) (Kovarik et al., 2008; Shaw and McKeown, 2011). Primers used to quantify *Nb18S, Nb5.8S* and two regions of *Nb25S* ribosomal genes are represented by

convergent arrows. **(B)** *Nb5S* rDNA is a pericentromeric region arranged in hundreds of tandemly repeated copies separated by intergenic spacers (IGS). Each copy contains the *Nb5S* ribosomal gene (Kovarik et al., 2008; Shaw and McKeown, 2011). Primers used to quantify *Nb5S* are represented by convergent arrows.

In the nucleus, actively transcribed rRNA genes are an integral part of the nucleolus, where the RNA polymerase I produces a long primary transcript (pre-ribosomal RNA *45S*), which is processed transcriptionally and post-transcriptionally to give the three rRNAs, *18S, 5.8S*, and *25S/28S*. The total maturation of the three rRNAs is completed in the cytoplasm (Henras et al., 2015). On the other hand, the precursor to the *5S* rRNA is synthesized by the RNA polymerase III. The transcriptional activity of these highly repetitive rRNA units is regulated by multilayered epigenetic mechanisms (Grummt and Pikaard, 2003; Tucker et al., 2010; Pontvianne et al., 2012; Dvorackova et al., 2015; Srivastava et al., 2016), but there is main difference between the *45S* and the *5S* units. While the transcription of the first one is controlled by the methylation of cytosine residues in the rDNA, the transcription of the latter (whose cytosines are also highly methylated) does not depend on DNA methylation (Mathieu et al., 2002).

The formation of extrachromosomal circular DNA (eccDNA) or episomal DNA, has been described in different organisms and they are primarily homologous to repetitive chromosomal DNA sequences, in particular those sequences that are organized in the genome as direct tandem repeats like histones or rDNA genes (Gaubatz, 1990). Genome instability of the rDNA has been observed in several organisms from yeast to humans including plants, and it could lead to the formation of episomal rDNA molecules (Cohen et al. 1997; Huang et al., 2008; Kobayashi, 2008; Srivastava et al., 2016). The constitutive transcription of the rDNA provokes the collision between the transcription and replication machineries that derives in changes in the copy number and the formation of episomal circular rDNA molecules (Dvorackova et al., 2015). The correct maintenance of the chromatin environment of the rDNA genes also plays and important role in their genomic stability. In mammalian double knockout cells for the CG maintenance methyltransferase DNMT1 (homologue to *A. thaliana* MET1) and *de novo* DNA methyltransferase DNMT3b (homologue to *A. thaliana* DRM1/DMR2), a significant increase of episomal rDNA was detected, probably due to intralocus recombination at the rDNA loci. In addition, the loss of CG methylation changes the chromatin status of the *45S loci* and increases the fraction of transcribed rRNA genes (Gagnon-Kugler et al., 2009). In *Drosophila*, Cohen and colleagues (2003) demonstrated the intramolecular homologous recombination between tandem direct repeats is a favourite mechanism for the formation of episomal circular rDNA molecules.

In Chapter 2, we reported that the *N. benthamiana 25S* ribosomal gene (*Nb25S*), a reference gene used to normalize DNA levels measured by quantitative PCR (qPCR) (Mason et al., 2007; Lozano-Durán et al., 2011; Rodríguez-Negrete et al., 2014), showed increased DNA levels in *NbMET1-* and *NbCMT3*-silenced plants. On the other hand, the DNA amount of other two non-tandemly-repeated *N. benthamiana* genes, such as *Actin (NbACT)* and *Elongation factor 1 α (NbEF1α)*, did not significantly differ between silenced and control plants. These findings suggested that the down-regulation of the *NbMET1* and *NbCMT3* DNA methyltransferases and their associated genome-wide hypomethylation, derived in the amplification of the *Nb45S* rDNA repeats, either by duplication of the repeats or by the formation of episomal circular molecules.

As this observation has not been previously described in plants, in this Chapter we have assessed the rDNA levels from the *45S* and the *5S* repeat units and the presence of episomal rDNA in *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants and in the *met1-3 A. thaliana* mutant.

RESULTS

3.1. Analysis of the ribosomal DNA (rDNA) levels in DNA methylation-deficient *N. benthamiana* plants.

In order to evaluate the impact of the DNA methylation machinery in the rDNA amount, VIGS was performed in *N. benthamiana* plants. Two fully expanded leaves of six-week-old plants were agroinfiltrated with either TRV-*NbMET1* or TRV-*NbCMT3* to induce the gene silencing of these two maintenance DNA methyltransferases (described in Chapter 2) or with the TRV vector as a control (mock plants). Naïve plants were also included in the analysis to rule out the possibility that presence of TRV could affect the amount of rDNA. To test the silencing degree of the DNA methyltransferases, total RNA was extracted from the second and third youngest leaves at 18 dpi, and the relative *NbMET1*, *NbCMT3* and TRV transcript levels were determined by RT-qPCR. We observed a relative transcript levels reduction of 81% and 77% for *NbMET1* and *NbCMT3*, respectively, when compared with TRV-treated plants (Figure 2A). While TRV RNA levels were similar between TRV-treated and TRV-*NbMET1*-inoculated plants, significant higher levels were found in plants that were agroinfiltrated with TRV-*NbCMT3* (Figure 2B). Total DNA was extracted from the same tissue and the relative DNA levels of the ribosomal genes from the *Nb45S* unit (*Nb18S, Nb5.8S* and two regions of *Nb25S*, named *Nb25S1* and *Nb25S2*) were quantified by qPCR, using *NbACT* and *NbEF1α* as normalizers. The primers used to amplify these *N. benthamiana* rDNA genes are depicted in Figure 1, and their sequences are listed in Table 2 from Materials and methods. The DNA quantification performed on these *NbMET1*- and *NbCMT3*-silenced plants revealed higher DNA levels of *Nb25S, Nb18S* and *Nb5.8S* when compared with TRV-treated plants (Figure 2C). In *NbMET1*-silenced plants, the increase was at least 4-fold for *Nb25S* and *Nb5.8S* and 2-fold for *Nb18S*, while the *NbCMT3*-silenced plants displayed a 2-fold increase for *Nb25S* and *Nb5.8S*, and 1.5-fold for *Nb18S* (Figure 2C). These results suggested that the silencing of the NbMET1 and NbCMT3 DNA methyltransferases and therefore the genome-wide hypomethylation this provokes (Figure 12 from Chapter 2), increases the amount of *Nb45S* rDNA in *N. benthamiana*. The significant higher accumulation of TRV detected in *NbCMT3*-silenced plants could not be responsible for this observation, as *NbMET1*-silenced plants showed lower TRV viral levels than *NbCMT3*-silenced plants and both displayed higher rDNA levels.
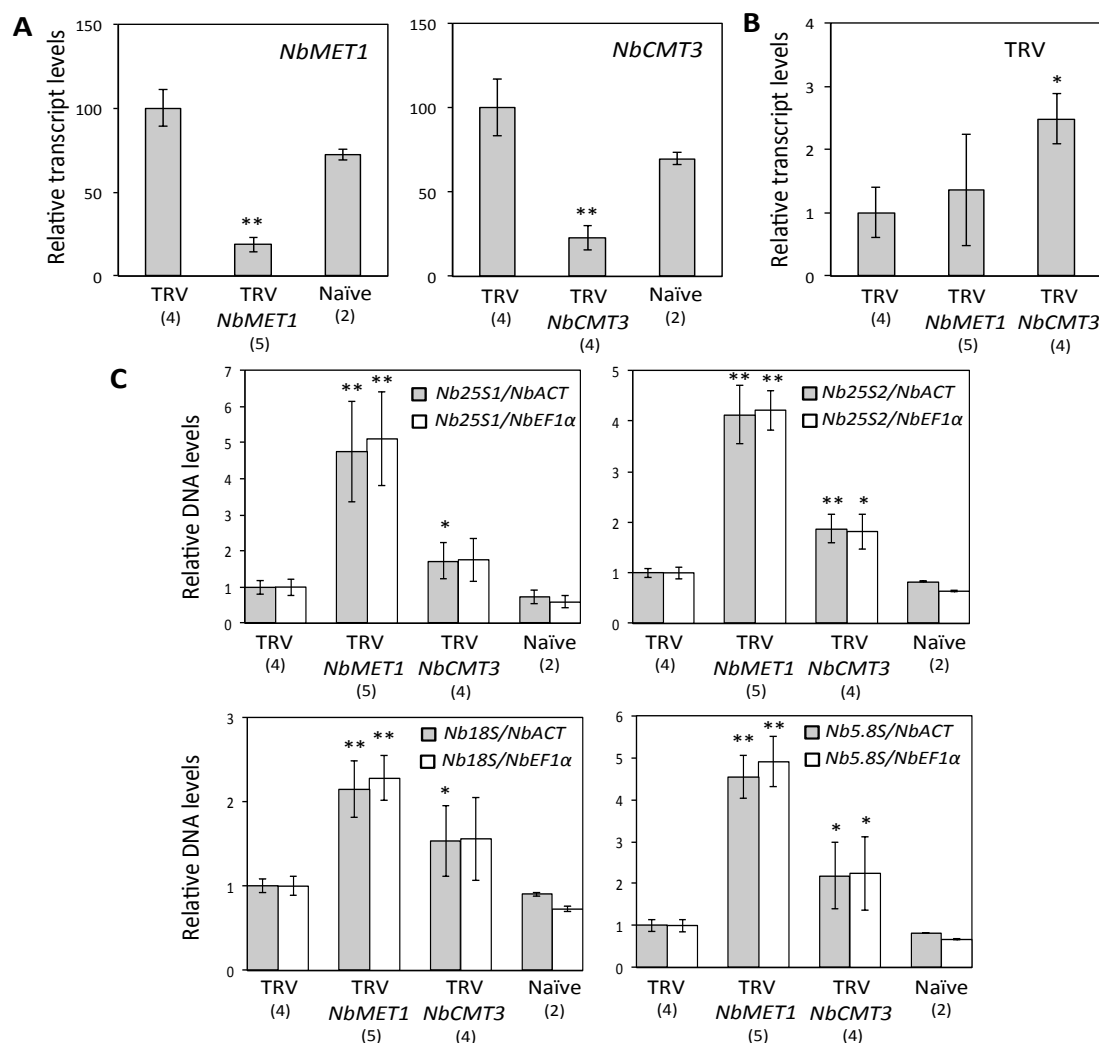
Figure 2. Relative transcript levels for *NbMET1* and *NbCMT3* genes and rDNA levels determined by qPCR in *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants. Two fully expanded leaves of six-week-old *N. benthamiana* plants were agroinfiltrated with TRV-*NbMET1,* TRV-*NbCMT3* and TRV (mock). At 18 dpi, total RNA was extracted from the pooled second and third most apical leaves and the relative transcript levels of **(A)** *NbMET1, NbCMT3* and **(B)** TRV were determined by RT-qPCR. Expression levels were normalized to *NbEF1α* and are presented as the relative expression compared with TRV-treated plants (set to 100 in **(A)** and set to 1 in **(B)**). **(C)** Total DNA was extracted from the same samples and the relative DNA levels of the ribosomal genes *Nb18S, Nb5.8S* and *Nb25S* (two regions named *Nb25S1* and *Nb25S2*) were determined by qPCR. DNA levels were normalized to *NbACT* (grey) and *NbEF1α* (white), and are presented as the relative levels compared with TRV plants (set to 1). Bars represent mean values ± standard deviation. Number of plants used is shown (n). Asterisks indicate significant differences according to Student's t-test (*, p-value < 0.05; **, p-value < 0.01).

We carried out a second biological replicate to confirm the rDNA levels increase in *N. benthamiana* plants whose maintenance DNA methylation machinery had been silenced. Additionally, we induced the silencing of *NbROS1* to assess if the impairment of the demethylation machinery has an effect on the rDNA levels. In order to address possible differences between both rDNA units, *Nb45S* and *Nb5S*, we also measured the rDNA levels of the *Nb5S* ribosomal gene. In silenced plants, a relative transcript levels reduction of 83%, 86% and 49% was observed for *NbMET1, NbCMT3* and *NbROS1*, respectively, when compared with TRV-treated plants (Figure 3A). No significant differences in TRV transcripts accumulation were

observed in TRV-treated and TRV-*NbMET1*- and TRV-*NbCMT3*-inoculated plants, but TRV-*NbROS1*-inoculated plants displayed significant lower levels (Figure 3B). In *NbMET1*-silenced plants, the rDNA levels increase was at least 3-fold for *Nb25S* and *Nb5.8S*, and 2-fold for *Nb18S*, while the *NbCMT3*-silenced plants showed a 1.5 to 2-fold increase for *Nb25S, Nb5.8S* and *Nb18S*, when compared with TRV-treated plants (Figure 3C). Conversely, we did not find significant differences as comparing the *Nb45S* DNA levels in *NbROS1*-silenced plants (Figure 3C). For the *Nb5S* gene, no significant differences in the rDNA levels were observed between the control plants (TRV-treated and naïve) and the silenced ones (Figure 3D).
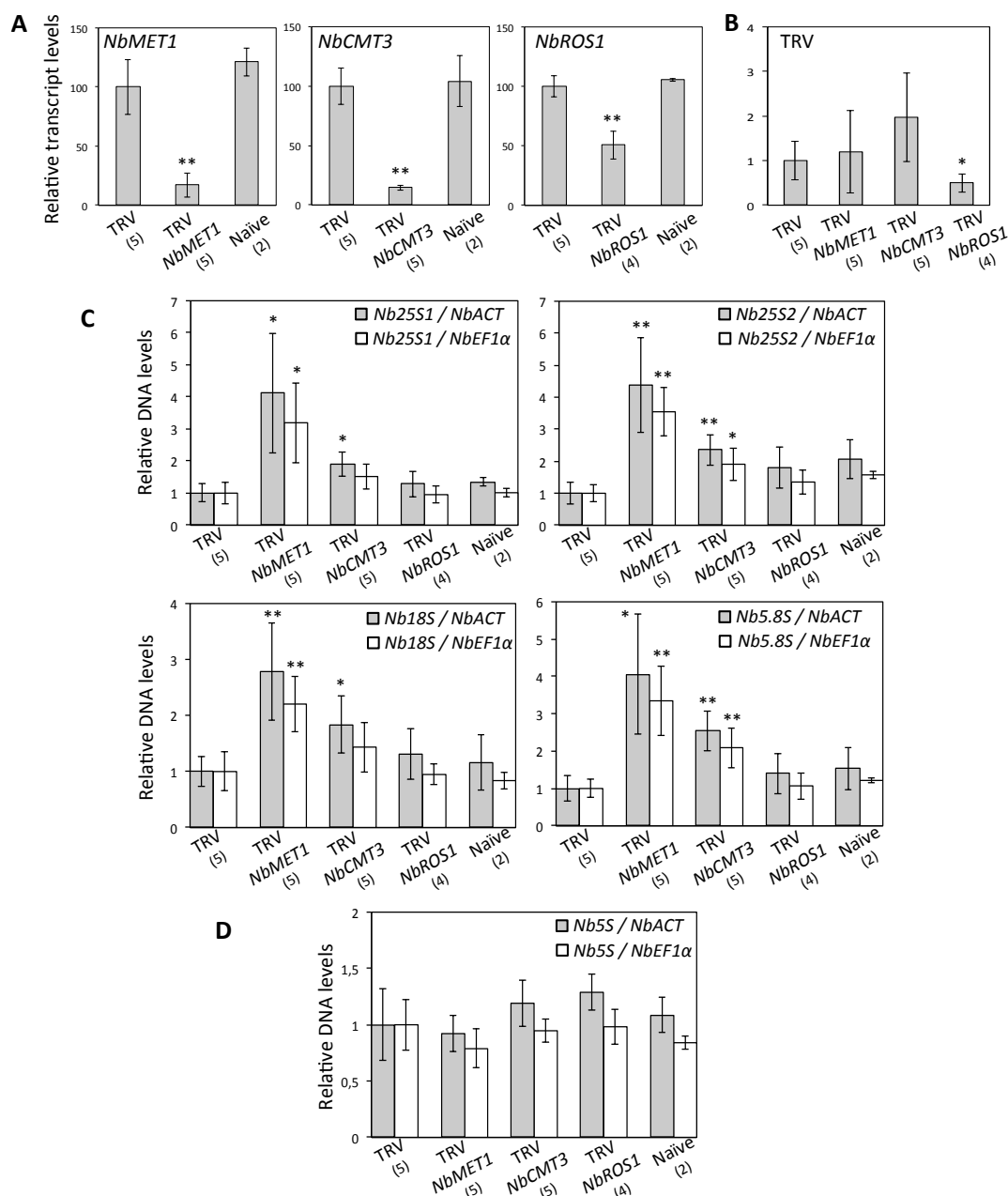

Figure 3. Relative transcript levels for *NbMET1*, *NbCMT3* and *NbROS1* genes and rDNA levels determined by qPCR in *NbMET1*-, *NbCMT3*- and *NbROS1*-silenced *N. benthamiana* plants. Two fully expanded leaves of six-week-old *N. benthamiana* plants were agroinfiltrated with TRV-*NbMET1,* TRV-*NbCMT3*, TRV-*NbROS1* and TRV as control. At 18 dpi, total RNA was extracted from the pooled second and third most apical leaves and the relative transcript levels

of **(A)** *NbMET1, NbCMT3, NbROS1* and **(B)** TRV were determined by RT-qPCR. Expression levels were normalized to *NbE1Fα* and are presented as the relative expression compared with TRV control plants (set to 100 in **(A)** and set to 1 in **(B)**). Total DNA was extracted from the same leaves than RNA and the relative DNA levels of the ribosomal genes **(C)** *Nb18S, Nb5.8S, Nb25S* (two regions called *Nb25S1* and *Nb25S2*) and **(D)** *Nb5S* were determined by qPCR. DNA levels were normalized to *NbACT* (grey) and *NbEF1α* (white), and are presented as the relative levels compared with TRV-treated plants (set to 1). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Asterisks indicate significant differences according to Student's t-test (*, p-value < 0.05; **, p-value < 0.01).

Together, these results suggested that the down-regulation of *NbMET1* and *NbCMT3*, but not *NbROS1*, induces an increase of the *Nb45S* rDNA amount. This effect occurred exclusively for the *45S* repeat unit, as no significant differences between the silenced and the TRV control plants were observed for the *Nb5S* repeat unit.

In order to determine if the down-regulation of the maintenance DNA methyltransferases also correlated with an increase in the rDNA levels in the model plant *A. thaliana*, we analysed the *45S* rDNA *locus* in a *met1-3* mutant (described in Chapter 2). Total DNA was extracted from pooled apical rosette leaves from four-week-old Col-0 and *met1-3* mutant plants (12 plants each). The relative DNA levels of *25S* and *18S* were determined by qPCR (primers alignment is shown in Figure 4A), and the *ACT2* gene was used as a normalizer. We observed a significant increase of the *25S* DNA levels in *met1-3* plants compared with Col-0, but no differences were detected for *18S* (Figure 4B). These results partially agree with those obtained in *N. benthamiana*, since we observed higher DNA levels for *Nb25S* and also *Nb18S* in *NbMET1*-silenced *N. benthamiana* plants. Our data suggested that the genome-wide hypomethylation induced by the reduced transcript levels of *MET1* (by VIGS in *N. benthamiana*, or in a *A. thaliana* mutant) have an impact on the rDNA amount, at least of the *25S* gene.
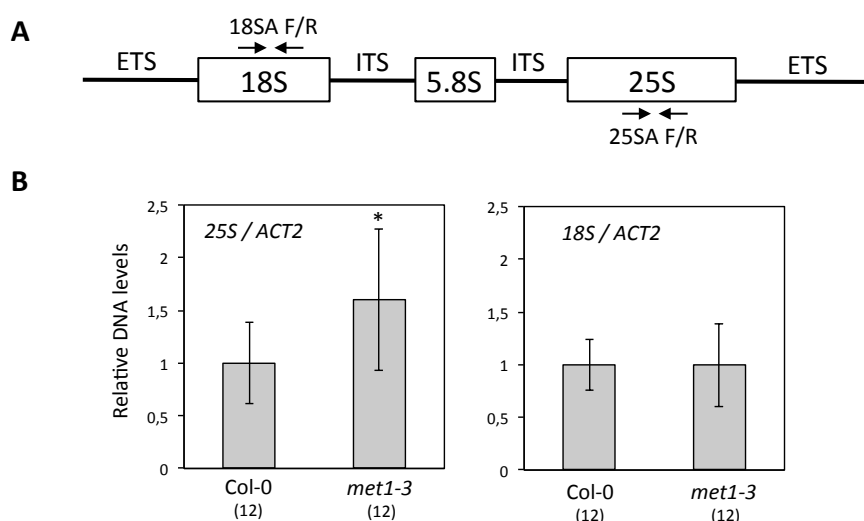

Figure 4. Relative rDNA levels determined by qPCR in *A. thaliana met1-3* mutant plants. (A) *A. thaliana 45S* rDNA diagram (not to scale) and alignment of specific primers (25SA F/R and 18SA F/R) for *25S* and *18S*. ITS: internal transcribed spacer. ETS: external transcribed spacer. **(B)** Total DNA was extracted from four pooled apical rosette leaves of four-weeks-old Col-0 and *met1-3* mutant *A. thaliana* plants (12 each). Relative DNA levels of *25S* and *18S* were determined by qPCR, using *ACT2* as a normalizer, and are presented as the relative levels compared with Col-0 plants (set to 1). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Asterisks indicate significant differences according to Student's t-test (*, p-value < 0.05).

3.2. Molecular mechanisms involved in the increase of *Nb45S* rDNA levels in DNA methylation-deficient *N. benthamiana* plants.

Together, our results indicated that in *N. benthamiana* and in *A. thaliana*, the down-regulation of the maintenance DNA methylation machinery and therefore, the genome-wide hypomethylation that this produces, induces an increase of the *45S* rDNA amount detected by qPCR. In order to confirm this observation by a different method, we determined the rDNA levels by Southern blot using a digoxigenin-labelled probe that corresponds to an 801 bp region of the *Nb25S* gene (Figure 5A). First, we assessed the technique capacity to detect this region and distinguish rDNA amount differences of 2- and 4-fold, since these were the increases observed using qPCR (Figure 2 and 3). Total DNA was extracted from naïve plants and several DNA amounts (10 µg, 5 µg and 2.5 µg) were *Eco*RI-digested and electrophoretically separated on an agarose gel (Figure 5B). As a control, 7 µg and 5 µg of undigested genomic DNA were loaded (Nd) and 10-fold dilutions of the *Nb25S* probe were also loaded on the same gel (10 ng, 1 ng, 0.1 ng and 0.01 ng). As represented in Figure 5A, the *Eco*RI digestion yields a 2608 bp fragment of the *Nb25S* gene that encompasses the *Nb25S* probe. Following the membrane hybridization with the *Nb25S* probe, two main bands were detected for lanes 3, 4 and 5 (digested DNA) (Figure 5C). The lower and more intense band corresponded to the 2608 bp *Nb25S* sequence, and showed differential signal intensity between lanes according to the total DNA amount that was loaded (Figure 5C). Therefore, we concluded that the Southern blot let us identify differences in the rDNA amount of around 2 to 4-fold.
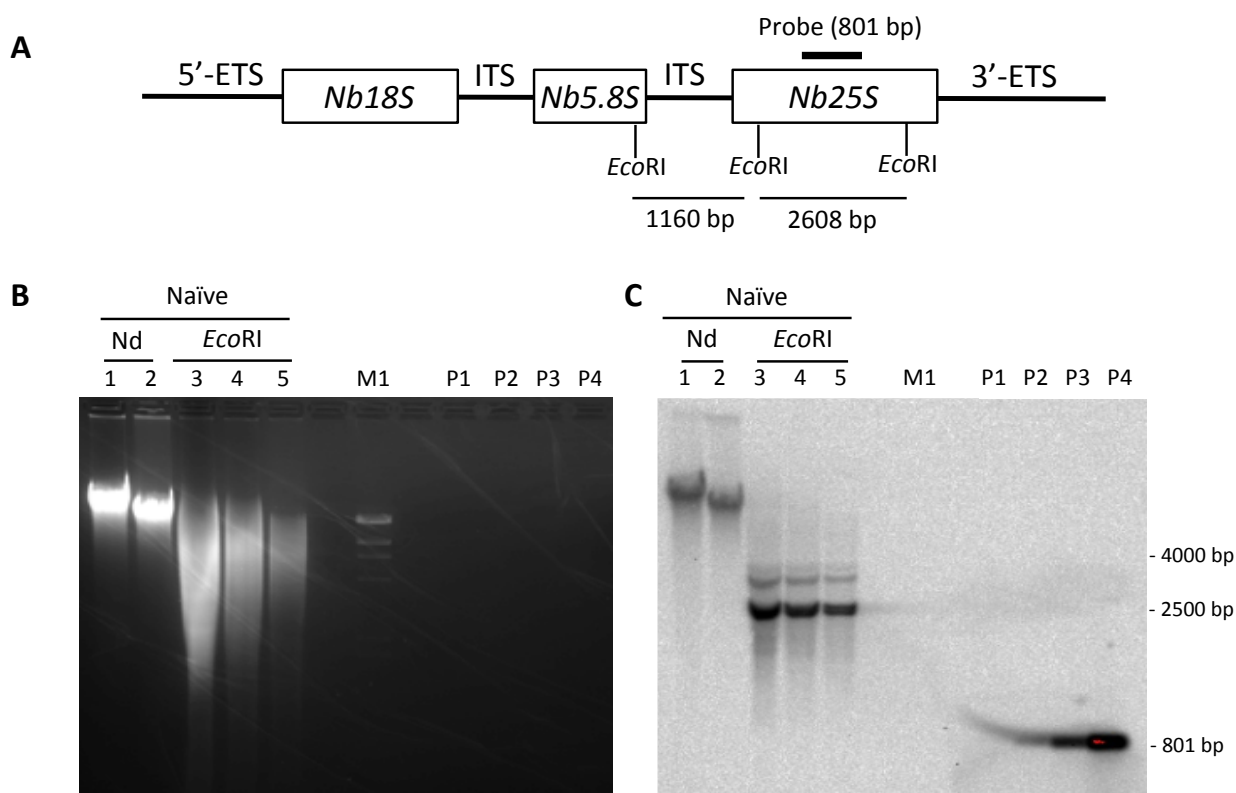

Figure 5. Southern blot analysis of total DNA from *N. benthamiana* naïve plants, digested with *Eco*RI and hybridized with a *Nb25S* probe. (A) Diagram (not to scale) of the *N. benthamiana Nb45S* rDNA. The *Eco*RI digestion sites and the location of the *Nb25S* probe are represented. **(B)** At 18 dpi, total DNA was extracted from the pooled

second and third most apical leaves of naïve plants, and 7 and 5 µg (lanes 1 and 2, respectively) of non-digested (Nd) DNA, and 10, 5 and 2.5 µg (lanes 3, 4 and 5, respectively) of *Eco*RI-digested DNA were loaded on a 1% agarose gel. **(C)** The membrane was hybridized with the *Nb25S* digoxigenin-labelled probe, which corresponds to an 801 bp region of *Nb25S*. The chemiluminescent signal was detected by using a luminescent imager system. P1, P2, P3 and P4 indicate 0.01, 0.1, 1 ng and 10 ng of DNA corresponding to the *Nb25S* probe, respectively. M: Lambda-*Hind*III, molecular marker.

Next, total DNA was extracted from the pooled second and third most apical leaves from naïve, TRV-treated and *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants, previously characterized and represented in Figure 3. For each DNA sample, 5 µg were digested with *Eco*RI and separated on an agarose gel (Figure 6A). The ethidium bromide gel was included as a loading control (Figure 6A). After the hybridization with the *Nb25S* probe, the same signal pattern as that shown in Figure 5B with a 2608 bp band for *Nb25S* and an unspecific upper band, was observed (Figure 6B). No significant differences in the band intensity among the different samples were observed. These results do not support the previous observation obtained by qPCR (Figure 2 and 3).
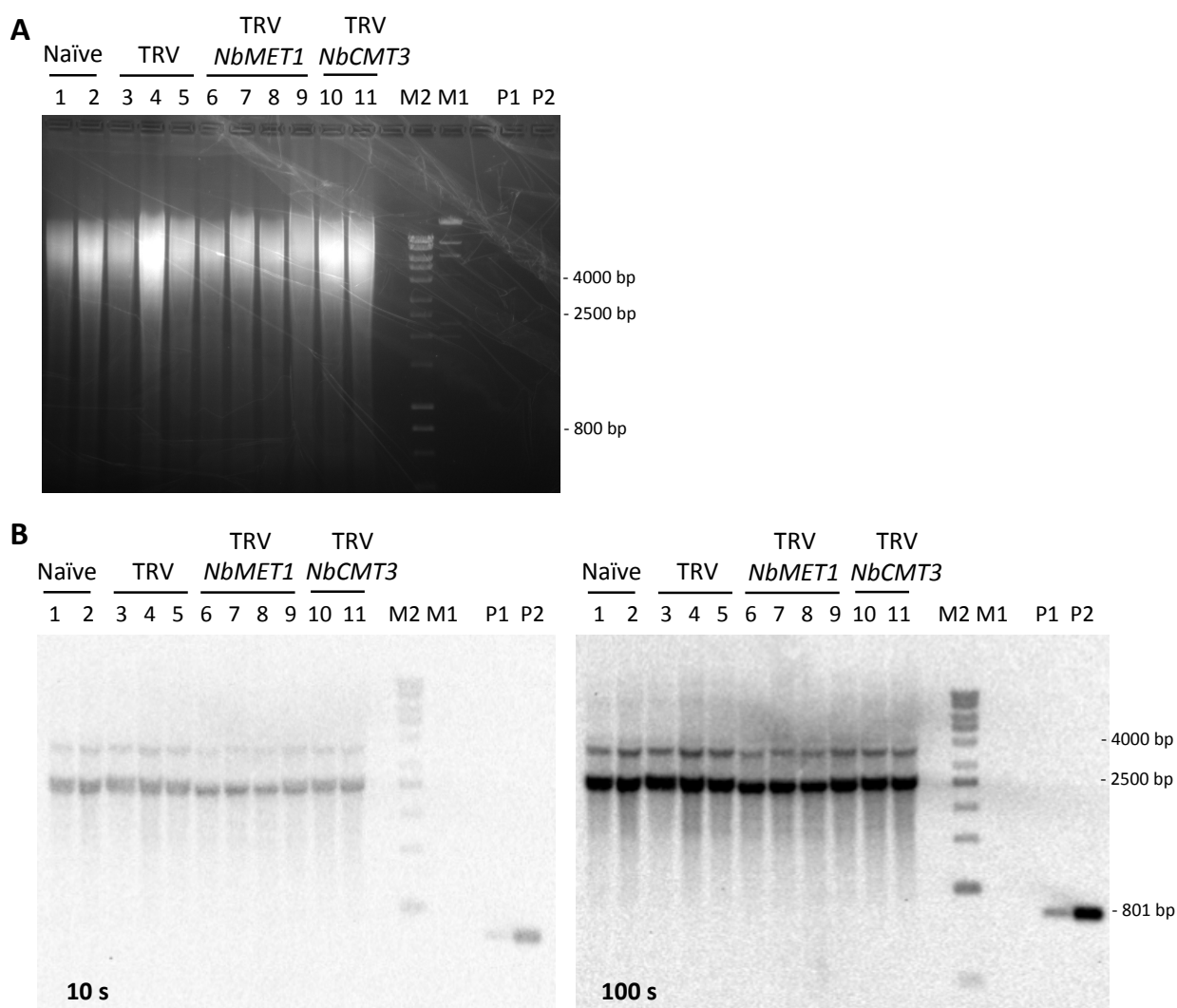

Figure 6. Southern blot analysis of total DNA from naïve, TRV-treated and *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants, digested with *Eco*RI and hybridized with the *Nb25S* probe. (A) At 18 dpi, total DNA was extracted from the pooled second and third most apical leaves of naïve, TRV-treated and *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants (previously characterized and represented in Figure 3), and 5 µg were digested with

*Eco*RI and loaded on a 1% agarose gel. **(B)** The membrane was hybridized with the *Nb25S* digoxigenin-labelled probe, which corresponds to an 801 bp region of *Nb25S*. The chemiluminescent signal was detected by using a luminescent imager system, at different exposition times (10 s and 100 s). Each number corresponds to a distinct sample (one plant). P1 and P2 indicate 0.1 ng and 1 ng of DNA corresponding to the *Nb25S* probe, respectively. HL: Hyperladder, molecular marker. M: Lambda-*Hind*III, molecular marker.

As previously mentioned, the presence of episomal rDNA molecules has been determined in *Drosophila*, yeasts and mammal cells, and their formation is thought to occur by mean intramolecular recombination between rDNA tandem copies (Cohen et al., 2003; Takeuchi et al., 2003; Kobayashi, 2008; Gagnon-Kugler et al., 2009). Based on these works, we hypothesized the potential formation of episomal circular molecules in the *Nb45S* rDNA *locus* by intramolecular recombination between the tandemly repeated copies in plants. As a first approach, we assessed the presence of episomal rDNA molecules in *N. benthamiana* by Southern blot using the *Nb25S* probe (Figure 5A). Total DNA was extracted from the pooled second and third most apical leaves from naïve, TRV-treated (negative control) and *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants (samples described in Figure 3), and electrophoretically separated the undigested DNA extraction on an agarose gel (Figure 7A). Following the membrane hybridization with the *Nb25S* probe, an intense band corresponding to the genomic DNA, and a smear, were observed (Figure 7B). We did not identify any smaller bands that could be attributed to episomal circular molecules (Figure 7B).
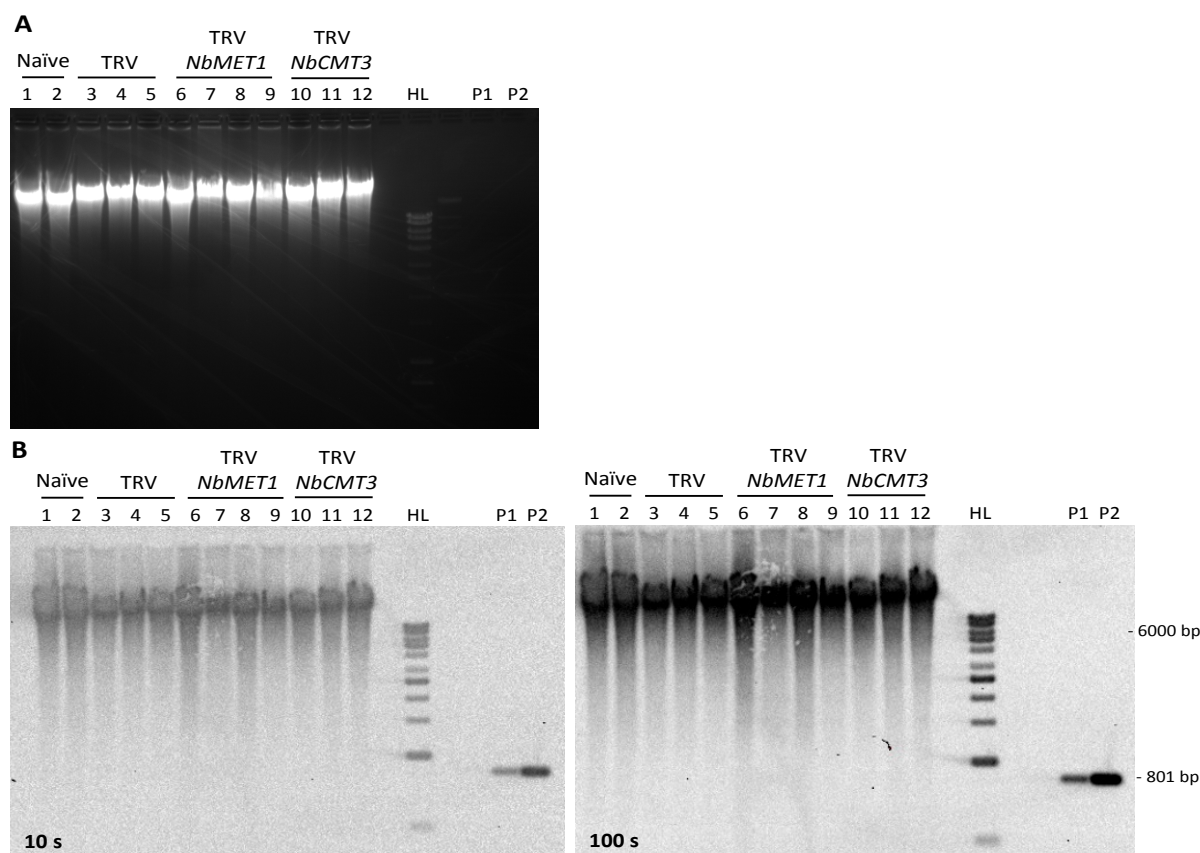

Figure 7. Southern blot analysis of total undigested DNA from TRV-treated and *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants, hybridized with the *Nb25S* probe. (A) At 18 dpi, total DNA was extracted from the pooled second and third most apical leaves from naïve, TRV-treated and *NbMET1*- and *NbCMT3*-silenced *N. benthamiana* plants (previously characterized and represented in Figure 3), and 5-8 µg were loaded on a 1% agarose gel. **(B)** The

membrane was hybridized with the *Nb25S* digoxigenin-labelled probe, which corresponds to an 801 bp region of *Nb25S*. The chemiluminescent signal was detected by using a luminescent imager system, at different exposition times (10 s and 100 s). Each number corresponds to a distinct sample (one plant). P1 and P2 indicate 0.1 ng and 1 ng of DNA corresponding to the *Nb25S* probe, respectively. HL: Hyperladder, molecular marker.

As a second approach, we evaluated the existence of episomal rDNA molecules by Rolling Circle Amplification (RCA), a method that uses the bacteriophage Phi29 DNA polymerase to exponentially amplify single- or double-stranded circular DNA templates through a rolling circle mechanism (Lizardi et al., 1998; Dean et al., 2001) (Figure 8).
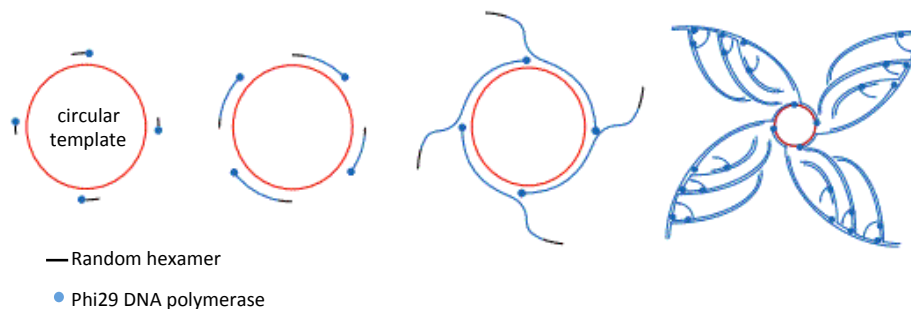

Figure 8. Scheme of the Rolling circle amplification (RCA) process. Random hexamer primers anneal to the circular template DNA at multiple sites and Phi29 DNA polymerase extends each of these primers. When the DNA polymerase reaches a downstream extended primer, strand displacement synthesis occurs. The displaced strand is rendered single-stranded and available to be primed by more hexamer primers. The process continues resulting in exponential amplification and the generation of concatemers (Lizardi et al., 1998; Dean et al., 2001).

Total DNA from TRV-treated and *NbMET1*- and *NbCMT3*-silenced plants (samples described in Figure 3 and 7) was amplified by RCA, and the product was digested with *Cla*I restriction enzyme, which cuts at a unique site within the *Nb45S* rDNA unit (Figure 9A).
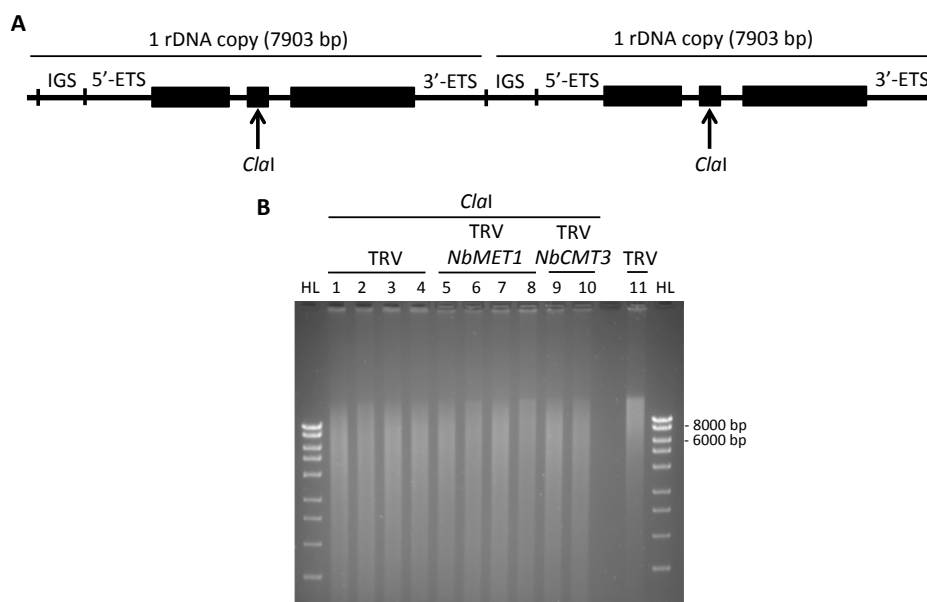

Figure 9. Electrophoresis of the *Cla*I-digested RCA product using as template DNA from TRV control and *NbMET1*- and *NbCMT3*-silenced plants. (A) Scheme of *45S* rDNA repeats. The *Cla*I restriction enzyme targets one site within each rDNA copy (7903 bp). **(B)** Total DNA from the TRV-silenced and TRV-treated plants represented in Figure 3 was RCA-amplified, digested with *Cla*I and loaded on a 1% agarose gel. Each number corresponds to a different sample (one plant). Number 11: non-digested sample. HL: Hyperladder.

According to our hypothesis, the RCA treatment will increase the amount of episomal rDNA molecules, and the digestion with *Cla*I will linearize these circular molecules generated by intramolecular recombination between different *Nb45S* rDNA copies (Figure 9A). As one *Nb45S* rDNA repeat contains 7903 bp, this band size would be detected after separating the digestion product on an agarose gel. However, we observed a smear corresponding to the *Cla*I-digested DNA, and no band between 6000 and 8000 bp was detected (Figure 9B).

The lack of a positive result prompted us to improve the capacity to detect episomal rDNA molecules by following a procedure that combined RCA and qPCR. First, we evaluated the approach feasibility using total DNA extracted from geminivirus-infected plants, since the viral DNA is a circular molecule. RCA was performed on total DNA samples from TYLCSV-infected plants, and this amplification was quantified by qPCR with specific primers for TYLCSV (hereafter TS, approach 1 in Figure 10). In parallel, the same amount of viral DNA from the same DNA samples was measured by qPCR without the previous RCA step (approach 2 in Figure 10). As indicated in Chapter 2, the Ct value derived from the qPCR analysis is inversely proportional to the original DNA amount of the gene of interest. We expected an increase in the TS amount after the RCA treatment and therefore, to find lower Ct values for RCA-treated DNA samples, compared to untreated DNA samples. On the other hand, the Ct values corresponding to a non-episomal genomic *locus* (*NbACT*) were expected to show a minimal Ct cycle variation regardless the RCA (negative control for RCA-qPCR) (Figure 10).
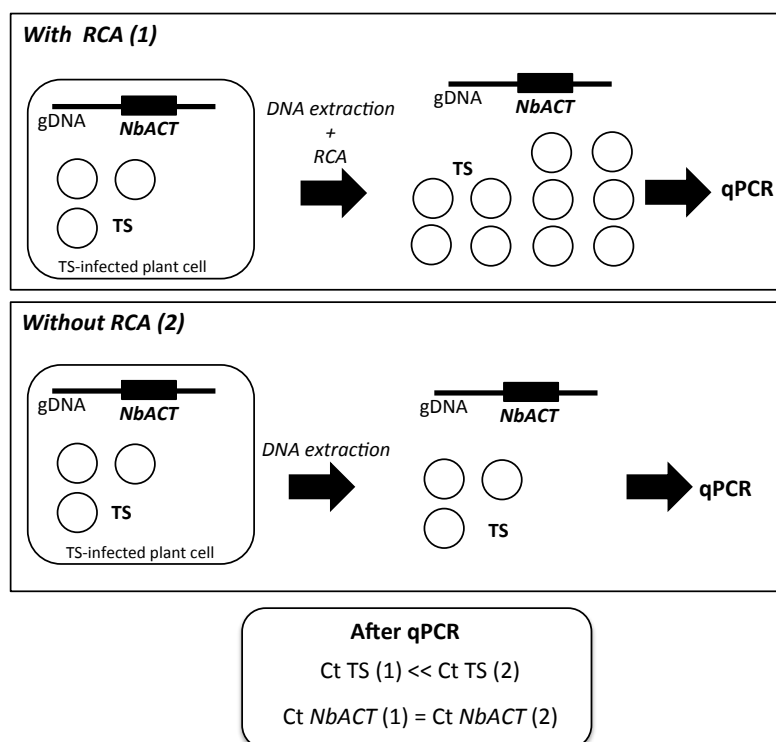

Figure 10. Workflow to determine the TS DNA levels by qPCR after RCA. Upper panel: DNA extracted from TS-infected *N. benthamiana* plants is used as RCA template and the product quantified by qPCR using TS and *NbACT* specific primers. Lower panel: DNA extracted from TS-infected plants is directly quantified by qPCR, without RCA step. Expected qPCR results show a TS Ct much lower when the RCA step is introduced (1), according to the increased TS amount. Minimal variation is expected for the *NbACT* Ct when comparing the conditions with (1) and without (2) RCA step. Black line represents genomic DNA. Circles represent TS DNA.

To accomplish the workflow depicted in Figure 10, total DNA from the second and third youngest leaves (pooled) of two mock-inoculated and two *N. benthamiana* plants infected with TYLCSV was extracted using two distinct methods. The first one (named Phenol method) is commonly used to measure viral accumulation in plants (see Materials and methods for details), while the second one (named Hirt method) is a modified protocol used in mammalian cells to enrich the total DNA extraction in episomal molecules (Hirt, 1967; Pradeep, 1993). Similar amounts of genomic DNA were observed in an agarose gel when comparing both extractions methods (Figure 11A). Subsequently, extracted DNA was RCA-amplified, and the product digested with *Eco*RI to yield linear molecules corresponding to the complete TYLCSV genome (2777 bp). After separating the RCA-digestion product on an agarose gel, bands with similar intensities that matched the TYLCSV size, were observed only in TYLCSV-infected plants (Figure 11B). These results indicate that RCA mostly amplified circular DNA corresponding to TYLCSV, resulting a similar amount of viral DNA regardless the extraction method performed.
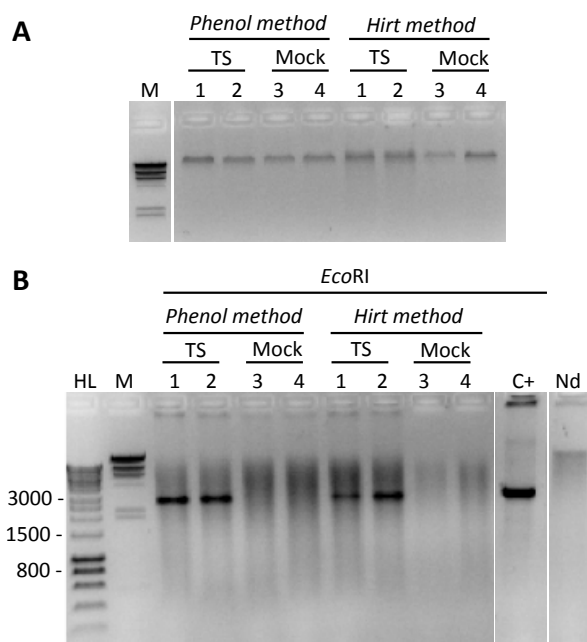

Figure 11. DNA accumulation of TS after RCA, in TS-infected and mock-inoculated *N. benthamiana* plants. (A) Agarose gel containing total DNA extracted from the pooled second and third youngest leaves of TS-infected and mock-inoculated *N. benthamiana* plants, by using the Phenol or the Hirt extraction method. **(B)** Total DNA shown in **(A)** was used as template for the RCA reaction and the product was digested with *Eco*RI and loaded on an agarose gel, yielding a band corresponding to the complete genome of TS (2777 bp). Each number corresponds to a different sample (one plant). C+: RCA product from pUC19 plasmid digested with *Eco*RI. Nd: Non-digested RCA product from the sample 2 extracted with the Phenol method. HL: Hyperladder, molecular marker. M: Lambda-*Hind*III, molecular marker.

Next, several dilutions (1:10, 1:100 and 1:1000) of the RCA product obtained from both TYLCSV-infected plants (TS1 and TS2, Figure 11), were analysed by qPCR with viral specific primers. As the extracted DNA was diluted 1:10 in the RCA reaction, the final dilutions were 1:100, 1:1000 and 1:10000. In parallel, DNA from the same samples but not RCA-treated (Figure 10), was diluted 1:100 and analysed by qPCR. In all dilutions assayed, the Ct values corresponding to the viral DNA (TS) resulted drastically reduced in RCA-treated samples as compared with the untreated ones (Figure 12A). Surprisingly, we observed a decrease (2 to 3 cycles) in the Ct value

for *NbACT* after RCA amplification (compared the 1:100 dilution of RCA-treated and untreated samples), but the Ct reduction was much pronounced (15 cycles aprox.) when the viral circular molecules were RCA-amplified (Figure 12A). When we measured the relative viral DNA amount using *NbACT* as a normalizer, we observed that the RCA-treated DNA samples contained much higher viral DNA levels than untreated samples (Figure 12B). These results demonstrate that the combination of RCA and qPCR improves the sensitivity to detect episomal circular molecules such as geminiviral DNA. Moreover, we conclude that the 1:10000 dilution is the most suitable to estimate the TS DNA levels in RCA-treated samples, as the 1:100 and 1:1000 dilutions yielded a TS Ct value that is not fully reliable (< 10). On the other hand, the Hirt method-extracted DNA samples contained lower TS DNA levels than the Phenol method-extracted samples, which suggests the Hirt procedure in plants did not produce an enrichment of circular episomal molecules (Figure 12B).

A

Phenol method

DNA sample	Dilution	Ct *NbACT*	Ct TS
Untreated TS1	1/100	25,49	17,15
RCA-treated TS1	1/100	22,94	1,72
RCA-treated TS1	1/1000	25,47	6,81
RCA-treated TS1	1/10000	28,38	11,34
Untreated TS2	1/100	24,89	17,82
RCA-treated TS2	1/100	23,24	3,18
RCA-treated TS2	1/1000	25,73	7,56
RCA-treated TS2	1/10000	28,89	11,89

Hirt method

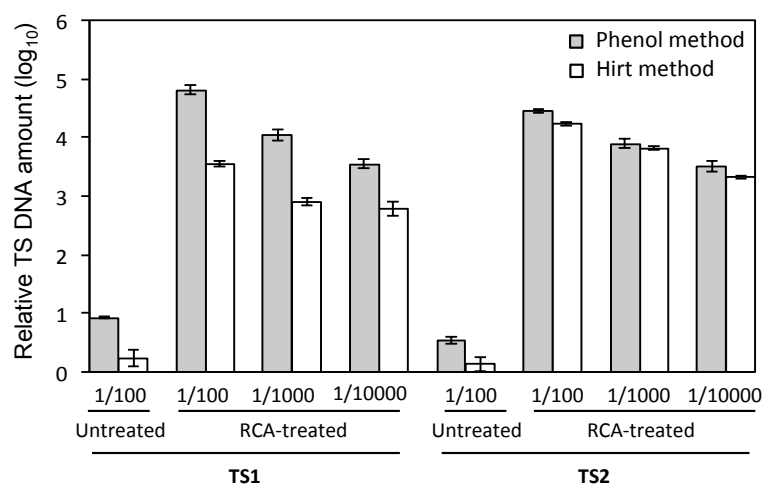| DNA sample	Dilution	Ct *NbACT*	Ct TS
Untreated TS1	1/100	24,92	18,89
RCA-treated TS1	1/100	22,53	5,49
RCA-treated TS1	1/1000	24,96	10,07
RCA-treated TS1	1/10000	27,89	13,38
Untreated TS2	1/100	25,54	19,81
RCA-treated TS2	1/100	22,9	3,57
RCA-treated TS2	1/1000	25,49	7,57
RCA-treated TS2	1/10000	28,26	11,95

B

Figure 12. Relative TS DNA levels determined by qPCR in RCA-treated and untreated TS-infected *N. benthamiana* plants. (A) Ct values corresponding to TS (viral gene) and *NbACT* (endogenous linear gene) were obtained by qPCR in the RCA-treated and untreated TS 1 and TS 2 DNA samples (extracted by the Phenol or the Hirt methods) represented in Figure 11. Several dilutions (1/100, 1/1000 and 1/10000) of RCA-treated DNA were used as qPCR template. **(B)** Relative TS DNA levels ($\log_{10}$) obtained by applying the Livak's data analysis to Ct values depicted in **(A)**. DNA levels were normalized to *NbACT*. Bars represent mean values ± measurement error of one plant**.**

As RCA coupled to qPCR is a reliable procedure to detect episomal circular molecules and increases its sensitivity, we used it to corroborate the existence of episomal rDNA molecules in TRV-*NbMET1* and TRV-*NbCMT3* silenced plants. We followed a similar workflow to that shown in

Figure 10, in which each DNA sample was split and this time, half of the sample was treated with the RCA kit containing the Phi29 DNA polymerase (Phi29 pol., approach 1 in Figure 13), and the other half with the same kit but lacking the Phi29 pol., which is similar to the RCA-untreated samples from Figure 10 (approach 2 in Figure 13). Then, the RCA product was analysed by qPCR using specific primers for *Nb25S* and *NbACT*. We expected to find a lower Ct value corresponding to *Nb25S* in Phi29 pol.-treated samples compared to the Phi29 pol.-untreated ones, while the *NbACT* Ct value was expected to show a low variation (Figure 13).
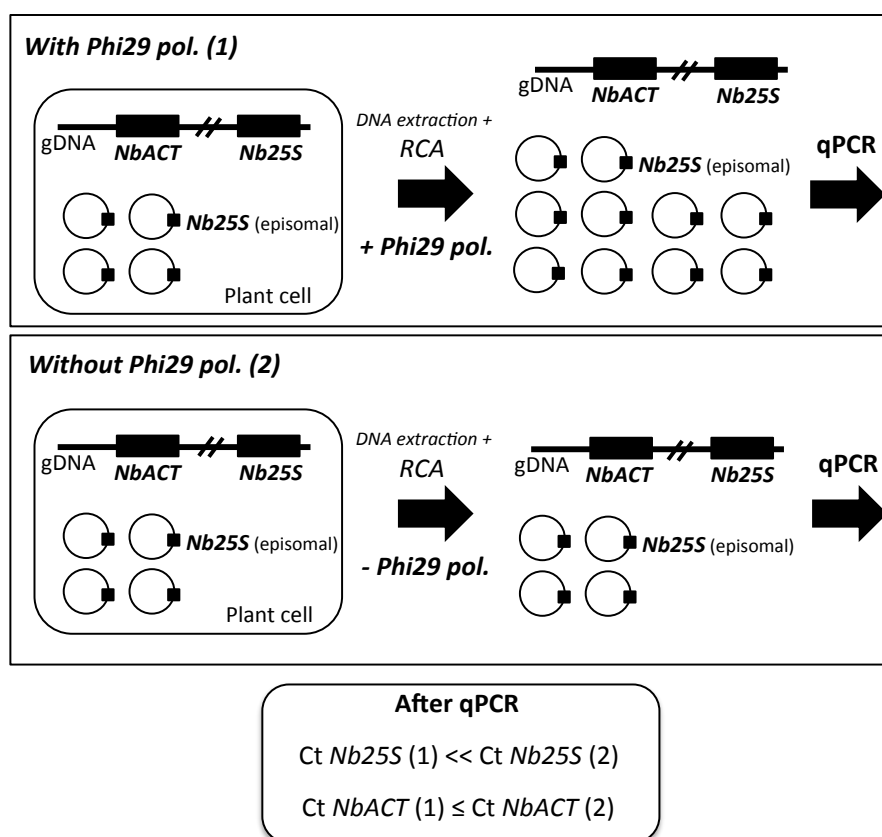

Figure 13. Workflow to determine the presence of episomal *Nb25S* levels by qPCR after RCA. Upper panel: DNA extracted from *N. benthamiana* plants is used as RCA template, and the product quantified by qPCR using *Nb25S* and *NbACT* specific primers. Lower panel: the same DNA is used as RCA template without addition of the Phi29 DNA polymerase, and the product is quantified by qPCR using *Nb25S* and *NbACT* specific primers. Expected qPCR results show an *Nb25S* Ct value much lower when the Phi29 pol. is added (1), according to the increased amount of hypothetical episomal rDNA produced by RCA. Conversely, a low variation is expected for the *NbACT* Ct value as comparing conditions With (1) and Without (2) Phi29 pol. Black line represents genomic DNA. Circles represent hypothetical episomal circular molecules of rDNA.

To accomplish the workflow depicted in Figure 13, first, total DNA from TRV-*NbMET1*- and TRV-*NbCMT3*-inoculated plants (samples described in Figure 3, 7 and 9) was RCA-treated with or without Phi29 pol., and the product was electrophoretically separated on an agarose gel. As expected, we only observed DNA amplification (visualized as a smear) when the polymerase was added (Figure 14).
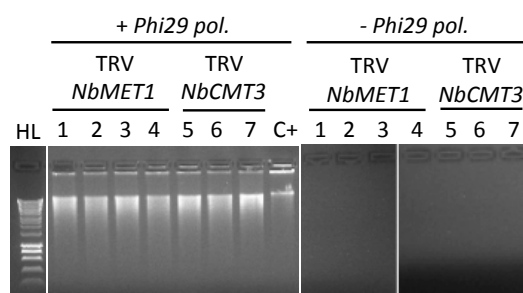
Figure 14. DNA accumulation after RCA with and without Phi29 DNA polymerase, in TRV-*NbMET1*- and TRV-*NbCMT3*-silenced *N. benthamiana* plants. Agarose gel containing the RCA product obtained with (+Phi29 pol.) and without (-Phi29 pol.) addition of the Phi29 DNA polymerase, using as template DNA extracted from TRV-*NbMET1*- and TRV-*NbCMT3*-inoculated plants (previously characterized in Figure 3, 7 and 9). Each number corresponds to a different sample (one plant). C+: RCA product from pUC19 plasmid. HL: Hyperladder, molecular marker.

Next, the RCA products (Figure 14) were diluted 1:100 and analysed by qPCR with specific primers for the region 1 of *Nb25S (Nb25S1)* (Figure 1) and for *NbACT*. When comparing the Ct values for *Nb25S1* for each sample subjected to RCA with or without Phi29 pol., a reduction of 4 cycles was observed, indicating that the RCA amplification had worked properly (Table 1). Surprisingly, a Ct reduction of 5 cycles was found for *NbACT* after the RCA amplification, suggesting that RCA was also amplifying a genomic *locus* that should not be in any episomal form (Table 1). These results suggest there is no accumulation of episomal molecules generated from the *Nb25S* ribosomal gene. The discrepancy between the decrease of the *NbACT* Ct value shown in Figure 12 (2-3 cyles) and in Table 1 (5 cycles), could be explained by the difference in the treatment of the DNA sample used as non-amplified control: RCA-untreated samples (Figure 12) or samples treated with the RCA kit lacking the Phi29 pol. (Table 1). The fact that these samples did not seem to actually have episomal DNA, as the TYLCSV-infected plants used to set the conditions for the RCA-qPCR approach had (Figure 10 and 12), could also increase the unspecificity of the RCA to amplify non-episomal DNA such as *NbACT*.

	Ct *Nb25S1* (Mean ± SD)		
Sample (n)	With *Phi29* pol	Without *Phi29* pol	Ct difference
TRV-*NbMET1* (3)	16.31 ± 0.57	20.49 ± 1.74	**4.18**
TRV-*NbCMT3* (3)	16.16 ± 0.34	20.45 ± 0.88	**4.29**

	Ct *NbACT* (Mean ± SD)		
Sample (n)	With *Phi29* pol	Without *Phi29* pol	Ct difference
TRV-*NbMET1* (3)	22.62 ± 0.53	27.92 ± 1.97	**5.3**
TRV-*NbCMT3* (3)	22.19 ± 0.19	27.46 ± 1.18	**5.27**

Table 1. Ct values of *Nb25S1* and *NbACT* obtained by qPCR after RCA (With and Without Phi29 DNA pol.) using DNA from TRV-*NbMET1*- and TRV-*NbCMT3*-silenced *N. benthamiana* plants. The RCA product (With and Without Phi29 DNA polymerase) from (n) TRV-*NbMET1* and TRV-*NbCMT3*-inoculated plants (Figure 14) was used as qPCR template with specific primers for *NbACT* and *Nb25S1*. Ct values (mean and standard deviation (SD)) corresponding to *NbACT* and *Nb25S1* are shown. One sample (TRV-*NbMET1* sample number 4 from Figure 14) yielded an *NbACT* Ct value higher than 31 and was not considered in the analysis.

3.3 Analysis of the *Nb45S* rDNA expression in DNA methylation-deficient *N. benthamiana* plants.

Although we were not able to confirm the increase in the amount of the *Nb45S* rDNA genes by other method than qPCR, we hypothesized that the transcription of the *Nb45S* rDNA locus could be induced in *NbMET1-* or *NbCMT3*-silenced *N. benthamiana* plants. As most of rDNA copies are transcriptionally silenced by DNA methylation and histone repressive modifications (Tucker et al., 2010; Pontvianne et al., 2012; Dvorackova et al., 2015; Srivastava et al., 2016), the DNA hypomethylation induced by the down-regulation of *NbMET1* or *NbCMT3* could activate their expression.

The RNA polymerase I transcribes the *Nb45S* rDNA genes as a precursor RNA (pre-rRNA) that is subsequently processed into the functional rRNAs *Nb18S, Nb5.8S* and *Nb25S* by the sequential deletion of internal and external transcribed spacers (ITS and ETS) (Henras et al., 2008) (Figure 15).
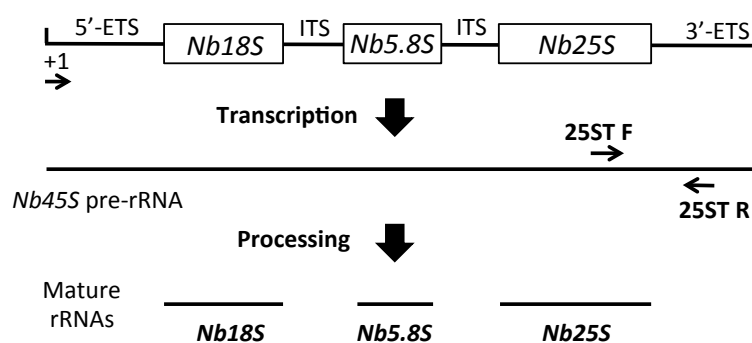

Figure 15. Diagram (not to scale) of the *N. benthamiana* Nb*45S* rRNA. The Nb*45S* rDNA is transcribed as a precursor rRNA (pre-rRNA) and subsequently processed into the mature rRNAs *Nb18S, Nb5.8S* and *Nb25S*, including deletion of the external and internal transcribed spacers (ETS and ITS). Primers used to measure the relative accumulation of pre-rRNA (depicted by arrows) are complementary to 3'-end of *Nb25S* (25ST-F primer) and the 3'-ETS region (25ST-R primer), flanking a 505 bp region.

To gain insight into the transcriptional control of the *Nb45S* rDNA genes by DNA methylation, we measured by RT-PCR the relative accumulation of the *45S* pre-rRNA in samples from TRV-treated and *NbMET1*-silenced plants (previously characterized and represented in Figure 3, 7 and 9, and Table 1). We focused on the *NbMET1*-silenced plants, instead of the *NbCMT3*-silenced ones, as they displayed higher increase of the rDNA levels when measured by qPCR. For this purpose, a one-step RT-PCR method was carried out (see Materials and methods) using different RNA amounts as template (400, 100 and 10 ng) and the same specific primers for the cDNA synthesis and the subsequent PCR (primers complementary to the 3'-end of *Nb25S* (25ST-F primer) and the 3'-ETS region (25ST-R primer) (primers alignment is shown in Figure 15). We followed the same procedure using primers for the *Rubisco* mRNA, which served as a control for the differences in the total RNA amount from each sample. As shown in Figure 16A, the *Rubisco* mRNA amount visualized on an agarose gel did not yield saturated signals in TRV-treated or *NbMET1*-silenced plants. For the *Nb45S* pre-rRNA, two main bands were amplified for each

sample, and just the upper one corresponded to the amplified region of the pre-rRNA (505 bp) (Figure 16B). No differences in band intensity for *Nb45S* or *Rubisco* were appreciated between TRV-treated and *NbMET1*-silenced plants and this result was confirmed by using the ImageJ software that measured the intensity of the bands shown in Figure 16A and 16B (Figure 16C). Therefore, we conclude that the relative *Nb45S* pre-rRNA accumulation was similar in TRV-treated and *NbMET1*-silenced plants.
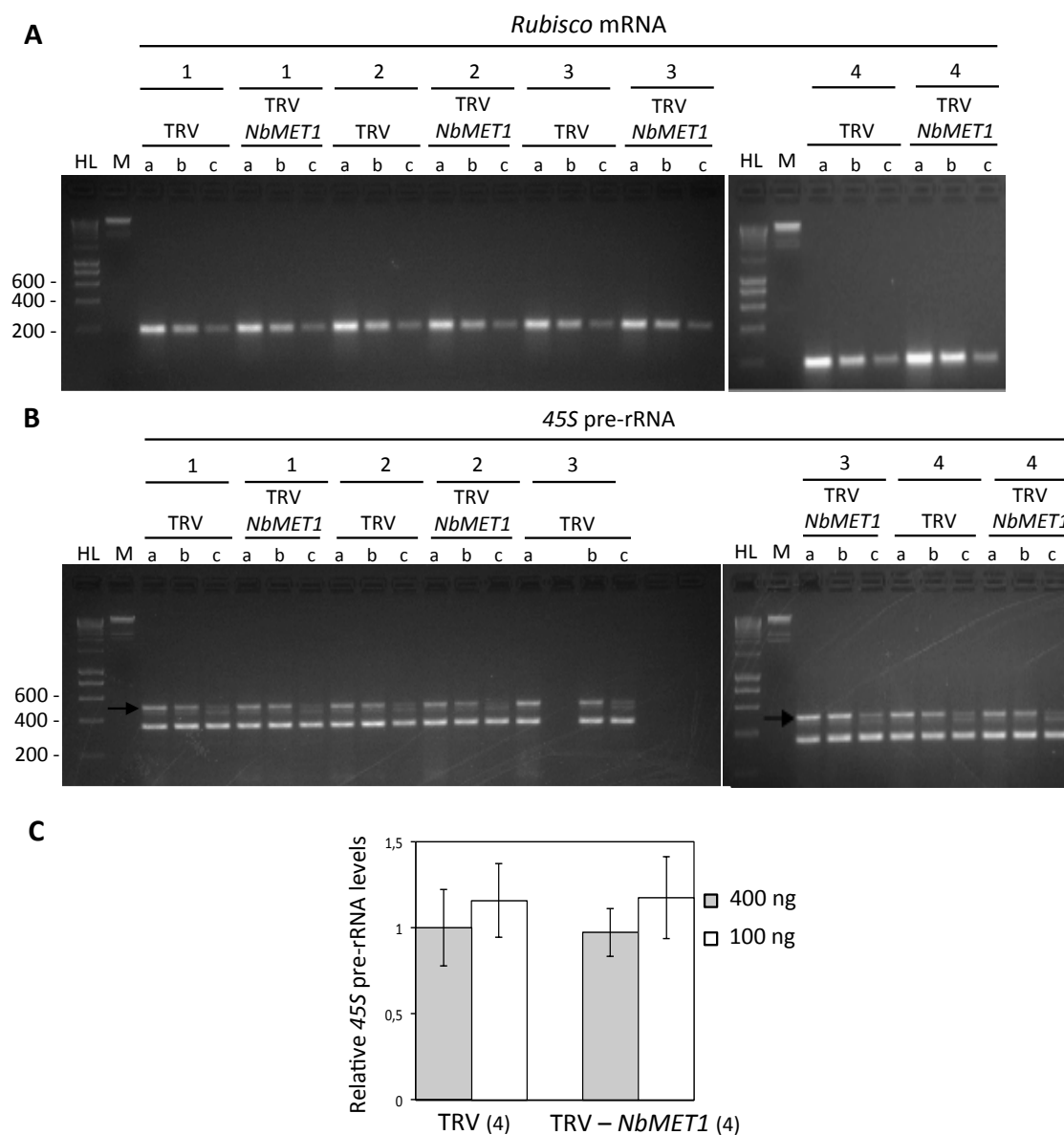

Figure 16. Relative *Nb45S* pre-rRNA accumulation determined by RT-PCR in TRV-treated and *NbMET1*-silenced *N. benthamiana* plants. Accumulation of **(A)** *Rubisco* mRNA and **(B)** *Nb45S* pre-rRNA, determined by RT-PCR, in RNA samples from four TRV- and TRV-*NbMET1*-inoculated plants (previously characterized and represented in Figure 3, 7 and 9, and Table 1). Numbers represent the RNA samples, and letters represent the RNA amount used as RT-PCR template (a: 400 ng; b: 100 ng; c: 10 ng). Arrows highlight the band corresponding to the amplified region of the *Nb45S* pre-rRNA (505 nt)**,** obtained with the 25ST-F and 25ST-R primers pair (see figure 15). Primers for *Rubisco* amplify a 184 nt region. HL: Hyperladder, molecular marker. M: Lambda-*Hind*III, molecular marker. **(C)** The relative *Nb45S* pre-rRNA accumulation normalized to the *Rubisco* mRNA was determined by measuring the band intensity with the ImageJ software. Results obtained with 400 ng (grey) and 100 ng (white) of template RNA are shown**.** Bars represent mean values ± standard deviation of (n) plants.

To rule out the possibility that the silencing of *NbMET1* could have an impact on the *45S* pre-rRNA processing and not on the transcription levels, we measured the relative transcript levels of the mature *Nb25S, Nb18S* and *Nb5.8S* rRNAs (see Figure 15) in the same RNA samples used to measure the *Nb45S* pre-rRNA. The cDNA synthesis was carried out using random primers, and the quantification was performed by qPCR using primers complementary to *Nb18S*, *Nb5.8S* and two regions of *Nb25S* (*Nb25S1* and *Nb25S2*) (same primers than used for the rDNA quantification, Figure 1). After the normalization to the *NbACT* transcript levels, the relative transcript levels of *Nb25S* and *Nb5.8S* resulted similar in TRV-treated and *NbMET1*-silenced plants. Surprisingly the *Nb18S* gene showed reduced transcript levels upon *NbMET1* down-regulation (Figure 17).
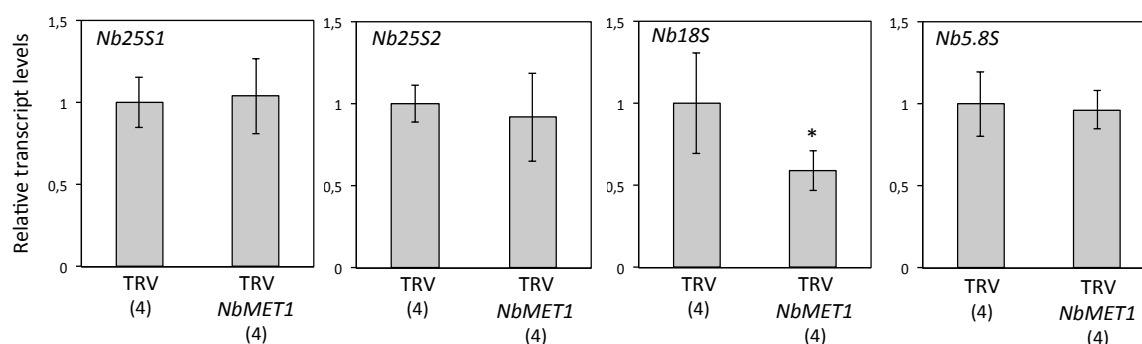

Figure 17. Relative accumulation of mature rRNAs determined by RT-qPCR in TRV-treated and *NbMET1*-silenced *N. benthamiana* plants. Relative transcript levels of the mature *Nb25S, Nb18S* and *Nb5.8S* rRNAs in TRV- and TRV-*NbMET1*-inoculated *N. benthamiana* plants, as determined by RT-qPCR. Transcript levels were normalized to *NbACT*, and are presented as the relative expression compared with TRV-treated plants (set to 1). Bars represent mean values ± standard deviation. The number of plants used is shown (n). Asterisks indicate significant differences as comparing TRV- and TRV-*NbMET1*-inoculated plants, according to Student's t-test (*, p-value < 0.05).

A similiar result was obtained when we measured the rRNA levels of *Nb25S* and *Nb18S* using a Bioanalyzer instrument, which allows an accurate RNA quantification without the cDNA synthesis step. This methodology is based on an automated electrophoresis to estimate the amount of total RNA, or of the mature *25S* and *18S* rRNAs, by determining the area under the RNA electropherogram. We analysed total RNA samples from TRV-treated and *NbMET1*-silenced plants represented in Figure 2, measuring two different concentrations (2 ng/μL (c) and the double, 4 ng/μL (2c)) for each RNA sample. Figure 18 shows the percentage of the total area, and thus the relative RNA amount, that corresponds to *Nb25S* and *Nb18S* for each individual sample (Figure 18A) or for the average (Figure 18B). No significant differences in percentage of area for *Nb25S* or *Nb18S* were observed between TRV-treated and *NbMET1*-silenced plants (Figure 18B), indicating that the accumulation of both rRNAs was similar in silenced and control plants. The percentage of area resulted lesser for *Nb18S,* compared with *Nb25S,* in most samples (Figure 18A), which highlights the RNA was not degraded. As expected, a similar area value was obtained when the two different concentrations (c or 2c) for each sample were measured (Figure 18).

Together, our observations indicated that the down-regulation of the NbMET1 activity does not induce the expression of the *Nb45S* precursor rRNA nor the processed *Nb25S, Nb18S* and *Nb5.8S* rRNAs**.
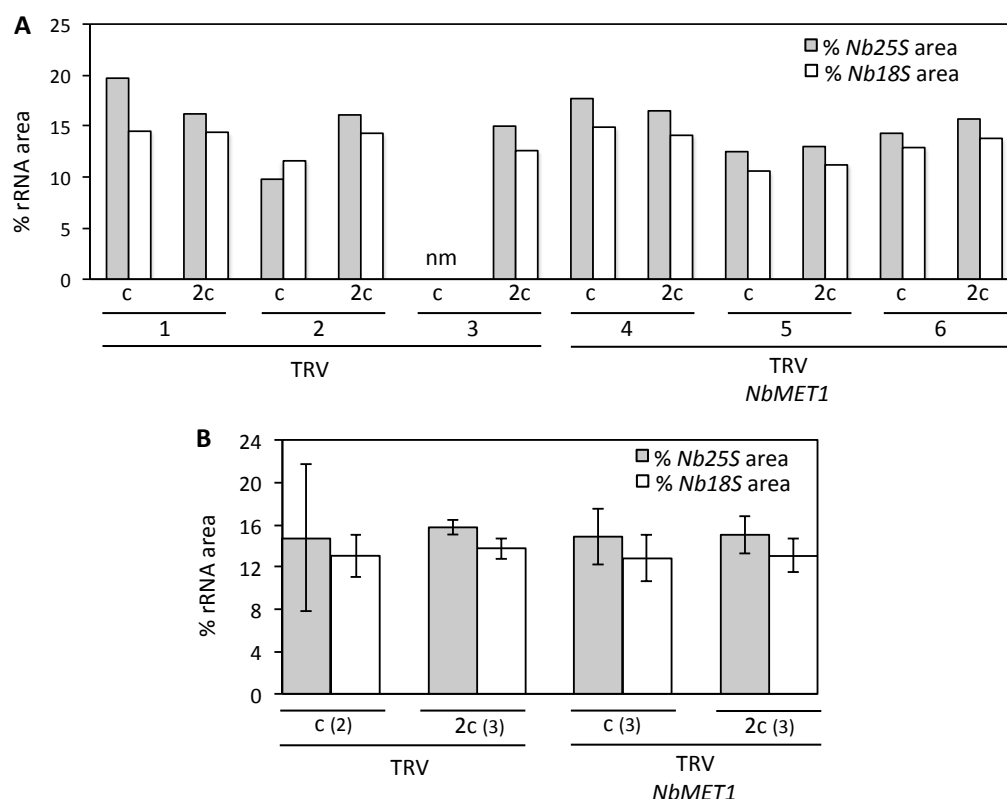

Figure 18. Relative accumulation of mature rRNAs determined by using a Bioanalyzer instrument in TRV control and *NbMET1*-silenced *N. benthamiana* plants. (A) Percentage (%) of the total area that corresponds to the *Nb25S* and *Nb18S* rRNAs, estimated in total RNA samples from TRV-treated and *NbMET1*-silenced plants. **(B)** Average of the percentages shown in **(A)**. Each number corresponds to one plant. For each sample, two concentrations (c and the double, 2c) were analysed. Bars represent mean values ± SD of (n) plants. Nm: not measured.

DISCUSSION.

The DNA levels of the *Nb25S, Nb18S, Nb5.8S* and *Nb5S* ribosomal genes were determined by qPCR in *N. benthamiana* plants down-regulated for *NbMET1*, *NbCMT3* or *NbROS1*. Using qPCR, we observed increased DNA levels for *Nb25S, Nb18S* and *Nb5.8S* genes (genes form the *Nb45S* rDNA unit) in those plants that displayed a reduction of around 70-80% in the transcript levels of the maintenance DNA methyltransferases, NbMET1 and NbCMT3. We just found this increase in the DNA levels for the *Nb45S* genes and not for the *Nb5S* rDNA gene, which is not unexpected as both rDNA repeated units showed several dissimilarities. Previously published data demonstrated that the transcriptional regulation, chromosome localization and chromatin

environment of both rDNA units are pretty different (Shaw and McKeown 2011, Henras et al., 2015). Also, the *met1-3 A. thaliana* mutant showed increased DNA levels for the *25S* rDNA gene but not for the *18S*.

Several works have demonstrated the existence of extrachromosomal circular DNA (eccDNA, episomal DNA) from yeast to humans. eccDNA is heterogeneous in size and composed of chromosomal sequences that contained the tandemly repeated genes such as rDNA genes. In *Drosophila,* distinct classes of circles derive from the rDNA cluster and are found throughout the fly's life cycle (Cohen et al., 2003). In *S. cerevisae*, protein complexes involved in DNA replication, DNA recombination and histone deacetylation, have a role in rDNA copy number regulation and in the formation of extrachromosomal rDNA circles from the rDNA repeats (Kobayashi, 2008). The work from Gagnon-Kugler et al. (2009) describes that the inactivation in mammalian cells of the maintenance DNA methyltransferase DNMT1 (homologue to MET1) and *de novo* DNA methyltransferase DNMT3b (homologue to DRM1/DRM2), leads to an increase in episomal rDNA and enhances ribosomal DNA recombination. These observations led us to hypothesize about the possible existence of additional, probably episomal, copies of *45S* rDNA molecules in *N. benthamiana*. This could explain the increase in DNA levels for the rDNA genes that we detected by qPCR when *NbMET1* and *NbCMT3* were silenced by VIGS (end of Chapter 2 and this Chapter). Therefore, we tried by two different approaches, Southern blot and RCA combined with qPCR, to detect if episomal rDNA molecules accumulated in *N. benthamiana* plants down-regulated for the maintenance DNA methyltransferases. However, we did not detect an increase in the amount of the *45S* rDNA nor the formation of rDNA episomal molecules by these approaches. Although the qPCR coupled to RCA constituted a reliable procedure that increased the sensitivity to detect episomal circular DNA such as geminiviral DNA, it also gave a high background amplification for a non-tandemly-repeated gene such as *NbACT*. Together, these results suggest that the silencing of *NbMET1* and *NbCMT3* and the genome-wide hypomethylation induced, have a significant impact on the accuracy for rDNA quantification by qPCR, but it does not produce an increase of the rDNA amount.

As the work from Gagnon-Kugler et al. (2009) describes that the loss of CG methylation in the double knockout DNMT1/DNMT3b cells, induced changes on the chromatin status of the rDNA *45S loci* and increased the fraction of transcribed rDNA genes, we hypothesized that the expression of the *Nb45S locus* could be induced in *NbMET1*-silenced plants, since this *locus* in plants is also transcriptionally regulated by DNA methylation and histone modifications (Tucker et al., 2010; Pontvianne et al., 2012; Dvorackova et al., 2015; Srivastava et al., 2016). Moreover, some works described that the HSV viroid induced hypomethylation and transcriptional reactivation of the *Nb45S* rDNA (Martínez et al., 2014; Castellano et al., 2015; 2016), corroborating the link between rDNA methylation and its expression in plants**.** Therefore, the relative transcript levels of the *Nb45S* pre-rRNA and the mature *Nb25S, Nb18S* and *Nb5.8S* rRNAs were determined by RT-PCR, RT-qPCR or by using a Bioanalyzer instrument, in TRV-treated and *NbMET1*-silenced plants. Results showed similar transcript levels in control and silenced plants,

which indicated that the impairment of the NbMET1-dependent CG methylation by VIGS, is not enough to reactivate the rDNA expression in *N. benthamiana* plants.

Chapter 4

No evidence of seed transmissibility of *Beet curly top virus* (BCTV) and *Cabbage leaf curl virus* (CaLCuV) in *Arabidopsis thaliana*

BACKGROUND

It has been believed that geminiviruses are restricted to horizontal transmission between neighbouring plants (Kashina et al., 2003) by mean insect vectors. Strikingly, recent works reported that the begomoviruses *Sweet potato leaf curl virus* (SPLCV) and *Tomato yellow leaf curl virus – Israel* (TYLCV) are seed-transmissible in sweet potato and tomato plants, respectively (Kim et al., 2015; Kil et al., 2016). The transmission rate for SPLCV to seeds and seedlings from the offspring of SPLCV-infected sweet potato plants is around 15% (Kim et al., 2015), while the transmission rate to the offspring seedlings from tomato plants infected with TYLCV, reaches 80 % (Kil et al., 2016). On the other hand, Rosas-Díaz and collaborators (2017) did not observe the seed-transmissibility of TYLCV in *N. benthamiana*. These contradictory results suggest that the host and/or the growth conditions are determinant for the seed transmissibility of a certain virus, since the experiment with TYLCV in tomato was carried out in greenhouses or with field-samples (Kil et al., 2016), while growth chambers were used to analyse the vertical transmission of TYLCV in *N. benthamiana* (Rosas-Díaz et al., 2017).

The seed transmission from curtoviruses has been also reported. *Beet curly top virus* (BCTV) and *Beet curly top Iran virus* (BCTIV) were seed-transmitted through three generations of petunia plants, showing a transmission rate of 38.2-78% for BCTV and 8.8-18.5% for BCTIV (Anabestani et al., 2017). On the other hand, Rajabu et al. (2018) did not observe symptoms or viral accumulation in two generations from the offspring of BCTV-infected tomato plants.

The knowledge about the geminivirus seed transmissibility constitutes a key point to design suitable strategies to control the viral disease in the field and to deeply understand the virus life cycle and its implications. Several reports described that geminiviruses modify the host DNA methylation pattern of certain plant *loci* (Buchmann et al., 2009; Zhang et al., 2011; Rodríguez-Negrete et al., 2013), and we are interested in determining the transgenerational inheritance of those epigenetic changes (Grossniklauss et al., 2013). To address this question, it is necessary to determine if the geminivirus under analysis is seed-transmissible. Therefore, the objective of this Chapter is to gain insight into the geminivirus transmission via seeds of the curtovirus BCTV and the begomovirus *Cabbage leaf curl virus* (CaLCuV) in *A. thaliana* plants.

RESULTS

4.1. Geminivirus presence in *A. thaliana* floral tissue.

The presence of geminiviruses in *A. thaliana* floral tissue will be essential to allow the viral vertical transmission on this model organism. We have previously described that the begomovirus TYLCV and the curtovirus BCTV highly accumulate in *A. thaliana* inflorescences (Figure 3 from Chapter 2), and we wondered if the geminiviruses were able to replicate in the inflorescences. We took advantage of a transgenic line generated in our group that let us determine if BCTV could replicate in the floral tissue. As part of his undergraduate dissertation, Jorge Sánchez Mingorance performed several experiments to detect the replication and monitor the infection of the geminivirus BCTV in *A. thaliana* transgenic plants that harbour a GFP cassette flanked by two repeats of the BCTV IR (2IRGFP). This system was previously characterized in *N. benthamiana* transgenic plants that contain the 2IRGFP cassette to detect the replication of *Tomato yellow leaf curl Sardinia virus* (TYLCSV). The accumulation of GFP correlates with the presence and active replication of the geminivirus, since the viral protein Rep induces the formation of GFP replicons and therefore, there is an increase in GFP expression. Under UV light, the replication of the geminivirus within the different plant tissues is detected by the GFP fluorescence signal (Morilla et al., 2006; Lozano-Durán et al., 2011). To determine if BCTV could replicate in *A. thaliana*, floral and rosette leaf (upper, middle and lower) tissues corresponding to BCTV- and mock-inoculated plants were visualized under UV light at different time points (21, 26 and 28 days post-inoculation- dpi). The fluorescence signal was not detected in mock tissues and was higher and more consistent in floral tissue of BCTV-infected plants than in rosette leaves, as just a subtle fluorescence signal was detected only in the upper leaves of the rosette (Figure 1A and 1B). This observation for the curtovirus BCTV is consistent with the work from Ascencio-Ibáñez et al. (2008), which described a higher accumulation for the begomovirus CaLCuV in upper than lower rosette leaves of *A. thaliana* infected plants (Ascencio-Ibáñez et al., 2008) and in inflorescences (Ascencio-Ibáñez, personal communication).

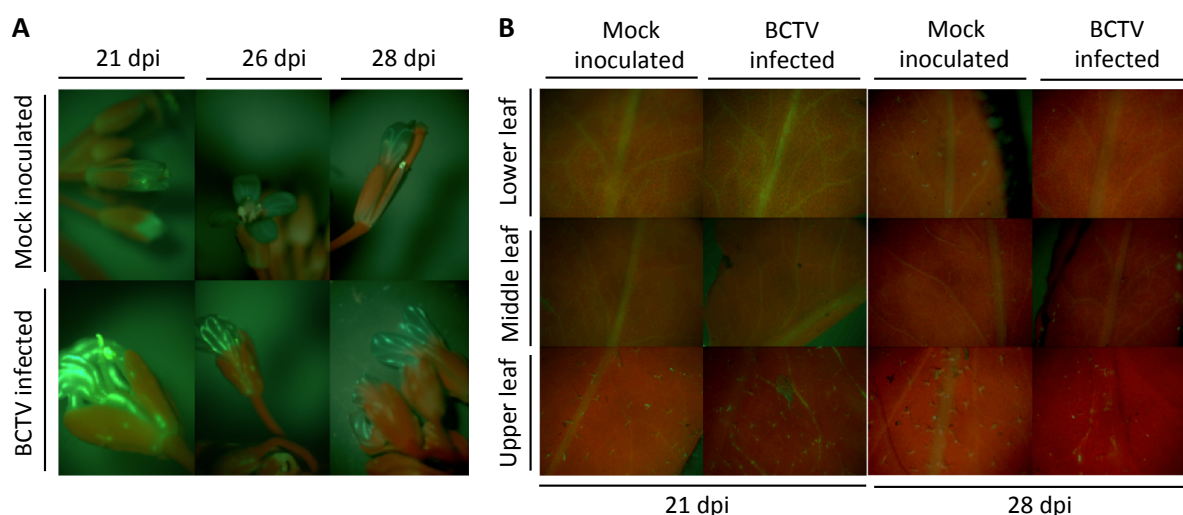**Figure 1. GFP-derived fluorescence in floral tissue and rosette leaves of BCTV-infected and mock-inoculated plants.** GFP-derived fluorescence observed under UV light in **(A)** floral tissue and **(B)** rosette leaves (upper, middle or lower) of BCTV-infected and mock-inoculated plants, at 21, 26 and 28 dpi. Representative images are shown (images taken from JSM´s undergraduate dissertation).

4.2. Selection of geminivirus-infected *A. thaliana* lineages.

To assess if geminiviruses are seed-transmissible in *A. thaliana*, we infected Col-0 plants with BCTV or CaLCuV, as both accumulate in floral tissue, reach high viral levels and produce severe symptoms during systemic infections (Ascencio-Ibáñez et al. 2008; Raja et al., 2008; Chapter 2 of this thesis). In two independent biological replicates, five-week-old *A. thaliana* plants were agroinfiltrated with the geminivirus or with the binary plasmid as a negative control (mock-inoculated). Three to four weeks after the agroinfiltration, total DNA was extracted from the three most apical rosette leaves and three inflorescences for each plant and the relative viral DNA levels were measured by qPCR on both tissues. In the first biological replicate (R1), the infection efficiency for BCTV and CaLCuV was 100% and 60%, respectively, and all infected plants produced seeds that were collected individually. In the second biological replicate (R2), the infection efficiency was 100% for both viruses, and for each virus, 4 out of the 5 infected plants produced seeds that were collected individually (Table 1). As expected, no viral DNA was detected in the mock-inoculated samples.

Biological replicate 1 (R1)			
Inoculated virus	Tissue-specific viral detection	Infected plants	Seeds production
BCTV	Leaves: 4 / 6	6 / 6	6 / 6
	Floral tissue: 6 / 6		
CaLCuV	Leaves: 1 / 5	3 / 5	5 / 5
	Floral tissue: 3 / 5		
Biological replicate 2 (R2)			
Inoculated virus	Tissue-specific viral detection	Infected plants	Seeds production
BCTV	Leaves: 4 / 5	5 / 5	4 / 5
	Floral tissue: 5 / 5		
CaLCuV	Leaves: 4 / 5	5 / 5	4 / 5
	Floral tissue: 5 / 5		

Table 1. Infection rate and seed production from *A. thaliana* plants infected with BCTV and CaLCuV. Two biological replicates were carried out with BCTV- and CaLCuV-inoculated *A. thaliana* plants (first column). The second column indicates the number of plants where viral DNA levels were detected in leaves or inflorescences. The third column indicates the number of infected plants out of the total inoculated plants. The fourth column indicates the number of plants that produced seeds.

To select the plants that displayed the highest BCTV and CaLCuV accumulation levels, we measured the viral DNA levels on the same qPCR experiment so we could compare the relative viral titers from infected plants from both biological replicates. Viral levels were measured from the floral tissue of the infected plants that were able to produce seeds (Figure 2). As a negative control, two mock-inoculated plants from each biological replicate were included. According to the results, a total of 5 plants for CaLCuV (1, 4, 5, 6 and 7) and for BCTV (1, 7, 8, 9 and 10) were selected to assess the vertical transmission of geminivirus.
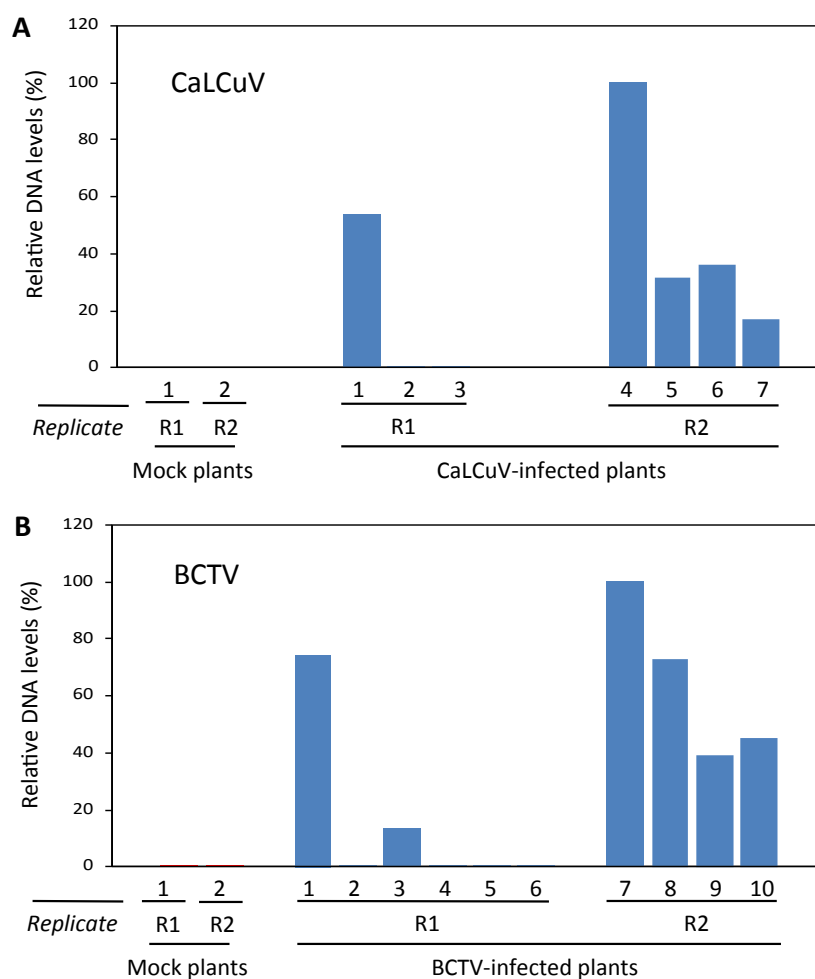
Figure 2. Relative viral DNA levels determined by qPCR in floral tissue of BCTV- and CaLCuV-infected *A. thaliana* plants. Total DNA was extracted from three pooled inflorescences of CaLCuV-, BCTV- and mock-inoculated plants. The relative DNA levels of **(A)** CaLCuV and **(B)** BCTV were determined by qPCR using *ACT2* as normalizer, and are presented as the relative accumulation compared with the highest value (set to 100%). Each bar and number represents one plant. The corresponding biological replicate (R1 or R2) is indicated (see Table 1).

4.3. Detection of geminivirus in the offspring from infected plants.

The viral transmission rate is the factor that determines the number of plants that need to be analysed in the offspring of an infected plant to conclude if a virus is or not seed-transmissible. This viral transmission rate may depend on the virus and on the host analysed and no data is so far available for the transmission rate of any geminivirus in *A. thaliana*. Therefore, we decided to analyse an extended number of offspring plants by making pools of plant material obtained from several infected plants and determine by qPCR the presence of BCTV or CaLCuV in the next generation. In order to decide the size of the pools for the offspring plants, first, we determined the sensitivity of the qPCR to detect one infected plant in a pool of an increasing number of non-infected plants. Three infected/uninfected ratios, 1/9, 1/24 and 1/49, were prepared from two independently infected plants per each geminivirus (BCTV1, BCTV2, CaLCuV1 and CaLCuV2) and two naïve (uninfected) plants. As positive and negative controls, we prepared

ratios of 1/0 and 0/1 respectively (Table 2). Equal amounts of DNA (2ng) were used for all the qPCR measurements.

DNA sample	Ratio BCTV-infected / Naïve	Ct BCTV
Naïve 1	0 / 1	30.34
BCTV 1	1 / 0	20.74
BCTV 1 + Naïve 1	1 / 9	24.8
BCTV 1 + Naïve 1	1 / 24	25.72
BCTV 1 + Naïve 1	1 / 49	26.95
Naïve 2	0 / 1	30.29
BCTV 2	1 / 0	21.86
BCTV 2 + Naïve 2	1 / 9	26.01
BCTV 2 + Naïve 2	1 / 24	26.64
BCTV 2 + Naïve 2	1 / 49	27.6

DNA sample	Ratio CaLCuV-infected / Naïve	Ct CaLCuV
Naïve 3	0 / 1	29.97
CaLCuV 1	1 / 0	19.91
CaLCuV 1 + Naïve 3	1 / 9	25.36
CaLCuV 1 + Naïve 3	1 / 24	26.37
CaLCuV 1 + Naïve 3	1 / 49	26.81
Naïve 4	0 / 1	30.96
CaLCuV 2	1 / 0	17.51
CaLCuV 2 + Naïve 4	1 / 9	21.66
CaLCuV 2 + Naïve 4	1 / 24	22.95
CaLCuV 2 + Naïve 4	1 / 49	23.95

Table 2. Ct values corresponding to different mixes of DNA from BCTV- or CaLCuV-infected and naïve plants. Total DNA from rosette leaves of two BCTV-infected plants (DNA samples: BCTV 1 and 2) or two CaLCuV-infected plants (DNA samples: CaLCuV 1 and 2) was mixed with total DNA from rosette leaves of different non-infected plants (DNA samples: naïve 1 to 4) (first column), according to different ratios showed in the second column (2 ng of the infected sample plus 18 ng of the naïve sample (ratio 1/9); 2 ng of the infected sample plus 48 ng of the naïve sample (ratio 1/24); and 2 ng of the infected sample plus 98 ng of the naïve sample (ratio 1/49)). DNA only from the infected sample was used as a positive control (ratio 1/0) and DNA only from the naïve sample as a negative control (ratio 0/1). All the mixes were diluted to 1 ng/μL, and 2μL (2ng) of each were used to determine the viral DNA levels by qPCR. The third column indicates the Ct value corresponding to BCTV or CaLCuV, obtained by qPCR.

Our data indicates that for both geminiviruses, we could detect 1 infected plant in a pool of 50, as the 1/49 ratio yielded a Ct value at least three cycles lower than the uninfected sample (ratio 0/1, DNA only from the naïve plant) (Table 2).

To assess the geminivirus vertical transmission in *A. thaliana*, seeds harvested from each of the selected infected plants (Figure 2), were grown for 25 days (all seedlings had developed at least, four true leaves) and 250 seedlings were sampled per each infected plant and bulked into pools of 50 seedlings (in total, 5 pools of 50 seedlings per infected plant) (Figure 3). Thus, 1250 seedlings belonging to 5 different plants or lineages, were analysed per each geminivirus. As a control, 150 seedlings coming from each mock-inoculated plant were collected and bulked into pools of 50 seedlings, and 3 mock plants were analysed (Figure 3). The offspring of each geminivirus was analysed separately in an independent experiment.
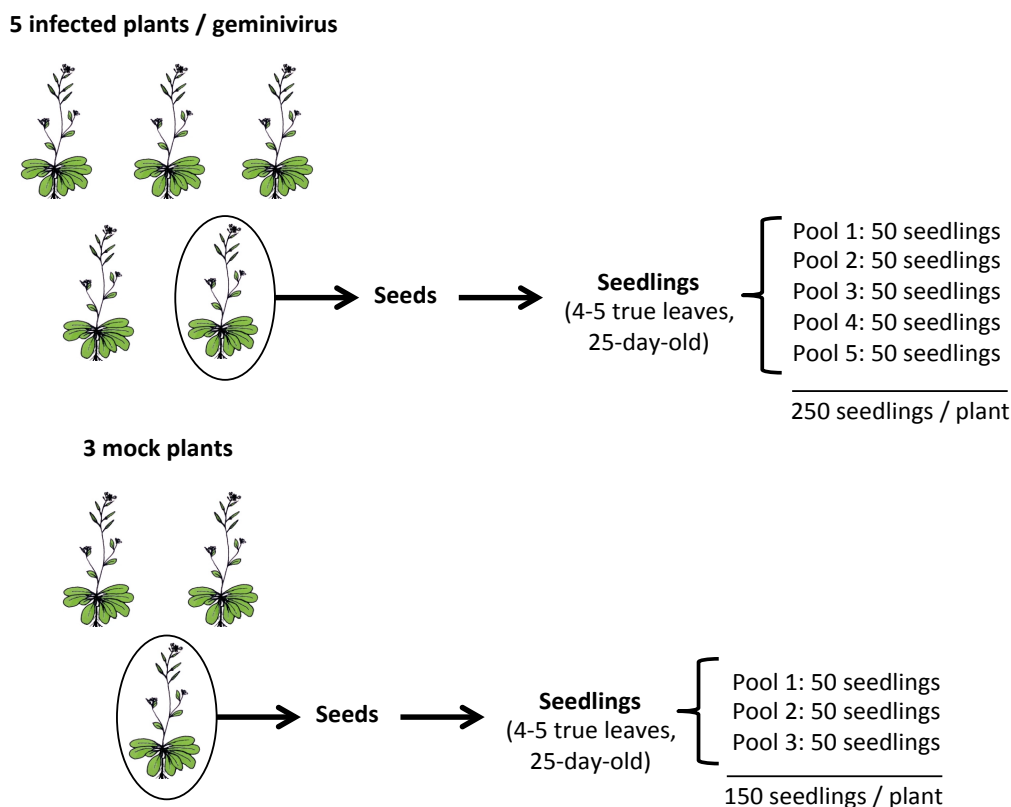
Figure 3. Experimental design to evaluate the seed transmissibility of BCTV and CaLCuV in *A. thaliana*. A total of 5 BCTV-infected plants and 5 CaLCuV-infected plants were selected to determine the geminiviral accumulation in the offspring. As a negative control, 3 mock-inoculated plants were introduced in the analysis. Seeds coming from each selected plant (infected or mock-inoculated) were grown for 25 days (seedlings contained 4-5 true leaves) and bulked into pools (samples) of 50 seedlings. A total of 5 pools per each infected plant (250 seedlings per each one) and 3 pools per each mock-inoculated plant (150 seedlings per each one) were generated.

Total DNA was extracted from the pools of seedlings coming from mock-inoculated, BCTV- and CaLCuV-infected plants (Figure 3) and the relative viral DNA levels were measured by qPCR using *ACT2* as normalizer. As a positive control, we measured the viral DNA levels in a CaLCuV- and a BCTV-infected plant. No significant differences were detected between the offspring of mock-inoculated and CaLCuV-infected plants (Figure 4A) indicating that we could not detect virus in any of these 1250 offspring plants and therefore there is no evidence of CaLCuV transmission via seeds under our experimental conditions.

On the other hand, some of the samples obtained from BCTV-infected plant showed a slightly higher signal using the primers to detect BCTV DNA, compared with the offspring from mock-inoculated plants (around 10-fold more), suggesting that limited amounts of viral DNA could be present in those samples (BCTV DNA levels in a BCTV-infected plant resulted from 1,000- to 10,000-fold the viral levels found in the offspring pools coming from BCTV-infected plants, Figure 4B).
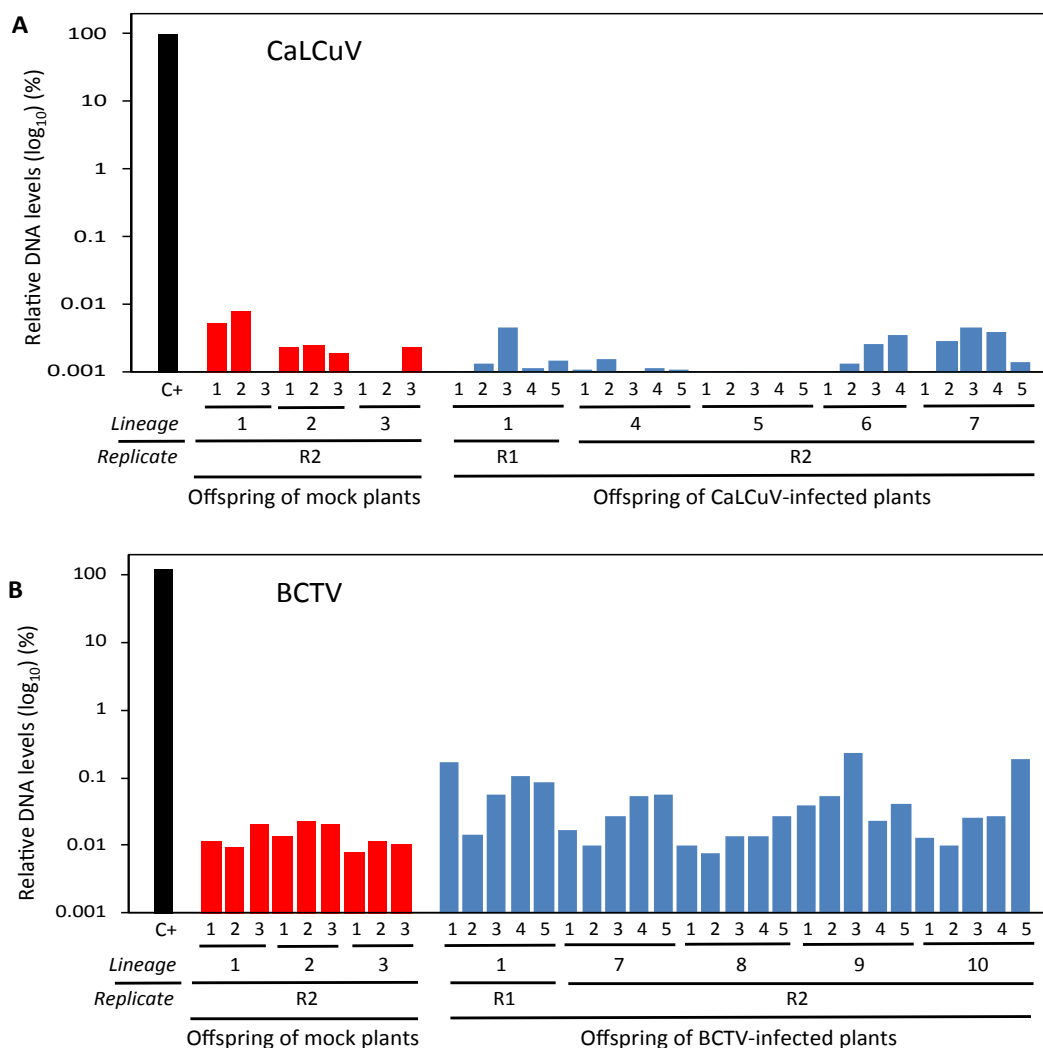
Figure 4. Relative BCTV and CaLCuV DNA levels determined by qPCR in the offspring of infected and mock-inoculated *A. thaliana* plants. Relative viral DNA accumulation in the pools of seedlings coming from **(A)** CaLCuV- or **(B)** BCTV-infected plants (blue), and mock-inoculated plants (red) (Figure 2 and 3), as measured by qPCR using *ACT2* as normalizer. The viral DNA levels are presented as the relative accumulation compared with the positive control (C+, set to 100%). Each value corresponds to one pool of seedlings (each number represents one pool). The lineage number corresponds to the plant number indicated in Figure 2. The corresponding biological replicate (R1 or R2) is indicated.

To determine whether the signal detected in those plants corresponded to the presence of residual amount of viral DNA, we analysed the samples using RCA (Rolling Circle Amplification) prior to the qPCR in order to increase the sensitivity of the viral detection. This approach increased the qPCR sensitivity to detect geminiviruses in infected samples in Chapter 3 (Figure 10 and 12). The RCA-qPCR was performed in the offspring of three BCTV-infected lineages, two that showed the highest BCTV levels in the previous qPCR (Figure 4B, lineages 1 and 9) and one that showed BCTV levels similar to those found in the offspring of mock plants (Figure 4B, lineage 8). DNA from the offspring of mock-inoculated plants was included in the analysis as a negative control, and as a positive one, DNA from a BCTV-infected plant was used. The high levels detected for BCTV in this positive control indicated that the RCA did amplify the circular

geminiviral DNA (1,000-fold compared to non-RCA treated infected sample, data not shown). No significant differences were found for the relative viral DNA levels between the offspring of BCTV-infected and mock plants (Figure 5). Therefore, we concluded that there is no evidence of the seed transmissibility of BCTV in *A. thaliana* under our experimental conditions.
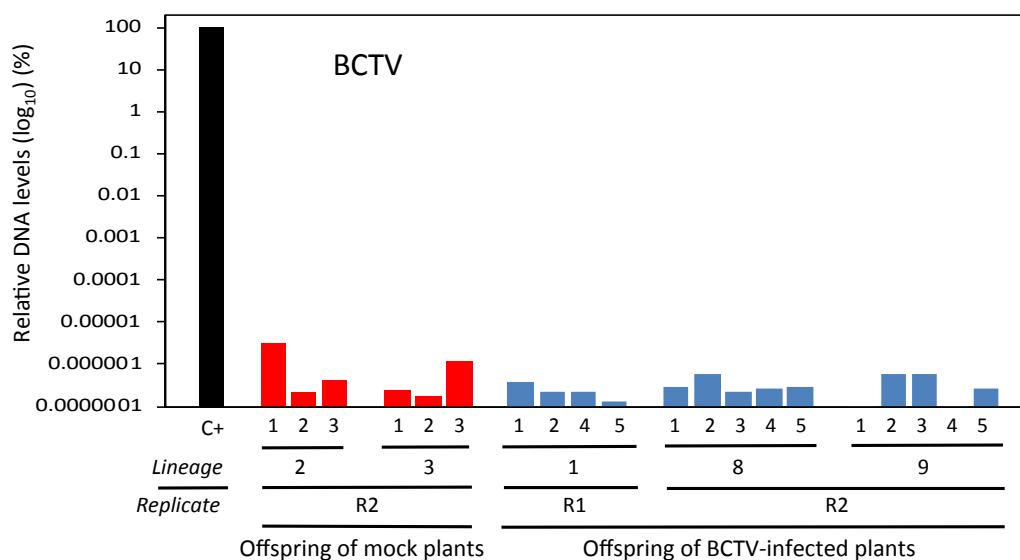

Figure 5. Relative BCTV DNA levels determined by qPCR following RCA. Total DNA from the pools of seedlings coming from BCTV-infected (blue) and mock-inoculated (red) plants (Figure 2 and 3) was amplified by RCA, and the relative BCTV DNA levels were determined by qPCR using *ACT2* as normalizer. The viral DNA levels are presented as the relative accumulation compared with the positive control (C+, set to 100%). Each value corresponds to one pool of seedlings (each number represents one pool). The lineage number corresponds to the plant number indicated in Figure 2. The corresponding biological replicate (R1 or R2) is indicated).

To corroborate that BCTV is not seed-transmissible in *A. thaliana*, we assessed the development of symptoms (stunted growth and atrophied inflorescences) in the offspring of BCTV-infected plants at 7 weeks post-germination. A total of 240 adult plants from BCTV-infected plants (from lineages 8 and 9, 120 plants each), and 80 plants from mock-inoculated plants, were visually analysed. No symptoms were observed in any of the offspring plants, but they were detected in the six plants infected grown under the same environmental conditions (Figure 6). Therefore, the lack of BCTV disease symptoms in the offspring of BCTV-infected plants at 7 weeks post-germination, lets us conclude once more, that BCTV is not seed-transmissible in *A. thaliana*.
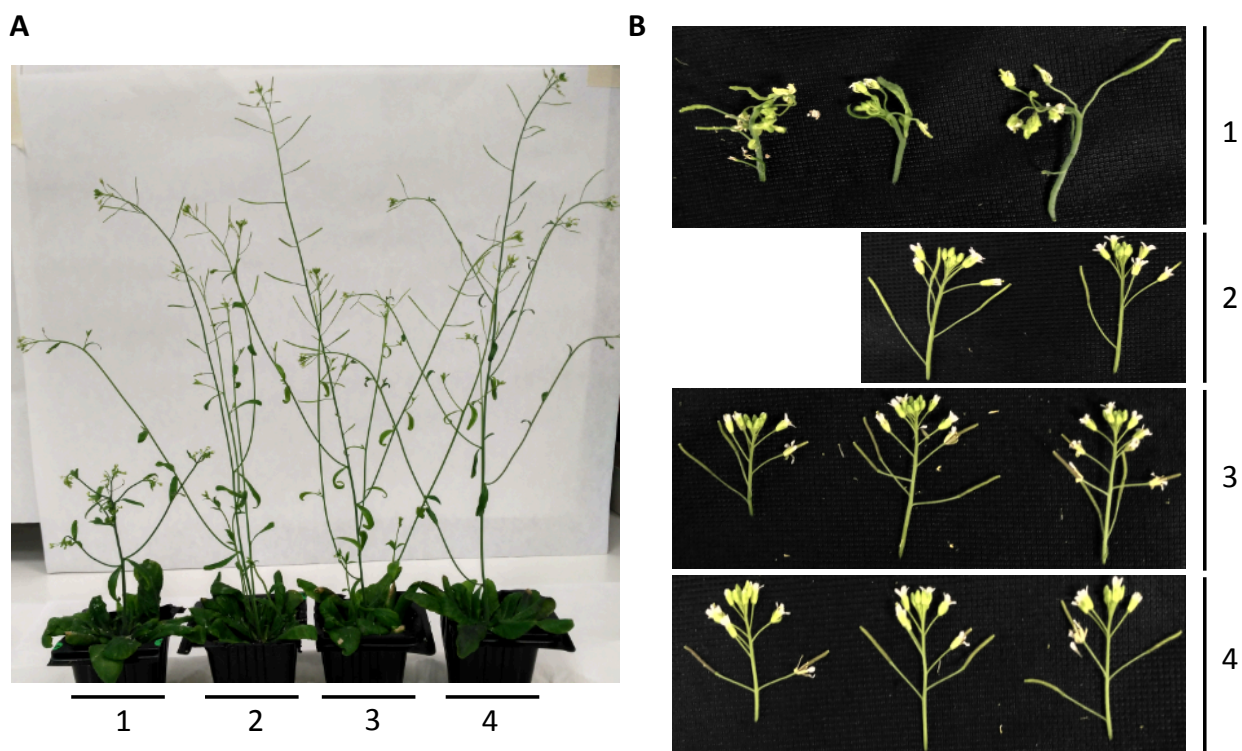
A

B

Figure 6. Analysis of BCTV disease symptoms in the offspring of BCTV-infected *A. thaliana* plants. A total of 240 and 80 plants coming from BCTV-infected and mock-inoculated plants, respectively, were grown for 7 weeks, and the development of BCTV disease symptoms was evaluated. **(A)** Complete view of four representative plants. **(B)** Detailed view of representative inflorescences. **(A, B)** 1: BCTV-inoculated plants coming from a BCTV-infected plant; 2: mock-inoculated plants coming from a BCTV-infected plant; 3: non-inoculated plants coming from a BCTV-infected plant; 4: non-inoculated plants coming from a mock-inoculated plant.

DISCUSSION

In this Chapter, we have studied the seed transmissibility of the curtovirus BCTV and the begomovirus CaLCuV in *A. thaliana*. We determined by qPCR the BCTV and CaLCuV accumulation in 25-day-old offspring seedlings from BCTV- or CaLCuV-infected plants, and could not detect the presence of these viruses. Moreover, for the offspring of BCTV-infected plants we performed a RCA-qPCR to increase sensitivity, and we assessed the development of BCTV disease symptoms in adult plants, with no positive results for BCTV seed-transmission. Based on our results we conclude that both geminiviruses are not seed-transmissible to the offspring in *A. thaliana* under our experimental conditions. Our results for BCTV differ to those presented by Anabestani and collaborators (2017) in petunia, where the infection rates in the offspring of BCTV-infected plants were greater than 38%. On the other hand, they agree with the work from Rajabu and collaborators (2018), who did not observe seed transmission of BCTV in tomato. Therefore, in

terms of seed transmissibility, BCTV seems to behave differently depending on the host. However, we cannot rule out the possibility that the growth conditions impact the efficiency of this process, since Anabestani and colleagues (2017) grew the plants in greenhouse, while we, and also Rajabu et al. (2018), performed the experiments in a growth chamber. Similarly, TYLCV was seed-transmitted in tomato plants grown in a greenhouse (Kil et al., 2016), but not in *N. benthamiana* plants grown in a growth chamber (Rosas-Díaz et al., 2017).

Our results have a potential impact on basic research, as investigators from the geminiviral community, especially those focused on studying transgenerational epigenetic inheritance, would find this observation relevant. Additionally, our work suggests that the seed-transmissibility of BCTV is not a general property of this virus, but depends on the host and/or growth conditions. This indicates that BCTV may be non-transmissible in other natural hosts, which is important to design control strategies in the field. To further investigate this topic we have to consider that other factors could affect the transmission of geminivirus to the next generation, such as the inoculation method or, as occurred in a natural infection, the presence of other viruses, pathogens or abiotic stresses.

CONCLUSIONS

1. TYLCV accumulation, gene expression and virus-derived small RNAs (vsRNAs) accumulation in tomato is similar regardless of the inoculation method, either by the bacteria *Agrobacterium tumefaciens* or by its natural vector, the whitefly *Bemisia tabaci*.

2. vsRNAs produced during the infection are mainly 21 and 22-nt long. They distribute along both strands of the entire TYLCV genome and accumulate in several large hotspot regions during infection in tomato plants. The increase in the vsRNA amount observed at 14 and 21 dpi is due to a rise of their redundancy.

3. In general, there is not a clear correlation between 24, 21 and 22-nt vsRNA hotspots and viral DNA methylation peaks, as there are viral regions that contained a high density of 24-nt vsRNA and just residual cytosine methylation, as well as other areas that contained high levels of DNA methylation and low amounts of 24, 21 and 22-nt vsRNA.

4. *A. thaliana* shows differential tissue-specific accumulation of geminiviruses, as the floral tissue sustains higher TYLCV, CaLCuV and BCTV DNA levels than rosette leaves.

5. The *A. thaliana* triple mutant *ddc* (*drm1-2 drm2-2 cmt3-11*), affected in the maintenance of methylation at CHG and CHH sites and the establishment of *de novo* methylation at CG, CHG and CHH sites, sustains slightly higher BCTV levels than Col-0 plants in floral tissue.

6. The lack of ROS1 (*de novo* demethylase), MET1 (responsible for the maintenance of methylation at CG context), or DRM1 DRM2 and CMT3 (*ddc* mutant) does not significantly impact TYLCV accumulation in *Arabidopsis thaliana*.

7. The down-regulation of *NbMET1*, *NbCMT3* or *NbROS1* by VIGS does not interfere with TYLCV or TYLCSV accumulation in *N. benthamiana*.

8. The biological importance of the plant DNA methylation machinery on geminivirus infection seems to be dependent on the host-virus analysed.

9. The down-regulation of *NbMET1* or *NbCMT3* induces an increase in the DNA levels of the ribosomal genes *25S, 18S* and *5.8S*, but not *5S*, in *N. benthamiana*. Such increase is detected by quantitative PCR (qPCR) but not by Southern blot or by RCA (Rolling Circle Amplification) coupled to qPCR.

10. There is no evidence of the presence of episomal rDNA molecules in non-silenced and *NbMET1-* or *NbCMT3*-silenced *N. benthamiana* plants.

11. The genome-wide hypomethylation produced by the down-regulation of *NbMET1* does not increase the expression of the precursor rRNA (pre-rRNA) or the rDNA genes in *N. benthamiana.*

12. The geminiviruses BCTV and CaLCuV are not seed-transmissible in *A. thaliana*.

References

Anabestani A, Behjatnia SAA, Izadpanah K, Tabein S, Accotto GP. 2017. Seed transmission of Beet curly top virus and Beet curly top Iran virus in a local cultivar of petunia in Iran. Viruses 9, 299-311.

Anfoka, G. Haj Ahmad F, Altaleb M, Al Shhab M. 2014. Detection of satellite DNA β in tomato plants with Tomato yellow leaf curl disease in Jordan. Plant Disease 98, 1017.

Aregger M, Borah BK, Seguin J, Rajeswaran R, Gubaeva EG, Zvereva AS, Windels D, Vazquez F, Blevins T, Farinelli L, et al. 2012. Primary and secondary siRNAs in geminivirus-induced gene silencing. PLoS Pathogens 8, e1002941.

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.

Ascencio-Ibáñez JT, Sozzani R, Lee TJ, Chu TM, 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 physiology 148, 436-454.

Baulcombe DC, Dean C. 2014. Epigenetic regulation in plant responses to the environment. Cold Spring Harbor Perspectives in Biology. doi, 10.1101/cshperspect.a019471.

Bian XY, Rasheed MS, Seemanpillai MJ, Ali Rezaian M. 2006. Analysis of silencing escape of Tomato leaf curl virus, an evaluation of the role of DNA methylation. Molecular Plant-Microbe Interactions 19, 614–624.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic, a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Bologna NG, Voinnet O. 2014. The diversity, biogenesis, and activities of endogenous silencing small RNAs in Arabidopsis. Annual Review of Plant Biology 65, 473–503.

Borah BK, Zarreen F, Baruah G, Dasgupta I. 2016. Insights into the control of geminiviral promoters. Virology 495, 101–111.

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

Briddon RW, Watts J, Markham PG, Stanley J. 1989. The coat protein of beet curly top virus is essential for infectivity. Virology 172, 628-633.

Briddon RW, 2009. Geminiviridae. John Wiley & Sons, Ltd.

Buchmann RC, Asad S, Wolf JN, Mohannath G, Bisaro DM. 2009. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. Journal of Virology 83, 5005–5013.

Calil IP, Fontes EPB. 2016. Plant immunity against viruses: antiviral immune receptors in focus. Annals of Botany 119, 711-723.

Campell BR, Song Y, Posch TE, Cullis CA, Town CD. 1992. Sequence and organization of 5S ribosomal RNA-encoding genes of Arabidopsis thaliana. Gene 15, 225-228.

Castellano M, Martínez G, Pallás V, Gómez G. 2015. Alterations in host DNA methylation in response to constitutive expression of Hop stunt viroid RNA in Nicotiana benthamiana plants. Plant Pathology 64, 1247-1257.

Castellano M, Martínez G, Marqués MC, Moreno-Romero J, Kohler C, Pallás V, Gómez G. 2016. Changes in the DNA methylation pattern of the host male gametophyte of viroid-infected cucumber plants. Journal of Experimental Botany 67, 5857-5868.

Castillo-González C, Liu X, Huang C, Zhao C, Ma Z, Hu T, Sun F, Zhou Y, Zhou X, Wang XJ, et al. 2015. Geminivirus-encoded TrAP suppressor inhibits the histone methyltransferase SUVH4/KYP to counter host defense. eLife, 1–31.

Ceniceros-Ojeda EA, Rodríguez-Negrete EA, Rivera-Bustamante RF. 2016. Two populations of viral minichromosomes are present in a Geminivirus-infected plant showing symptom remission (recovery). Journal of Virology 90, 3828–3838.

Chellappan P, Vanitharani R, Pita J, Fauquet CM. 2004. Short interfering RNA accumulation correlates with host recovery in DNA virus-infected hosts, and gene silencing targets specific viral sequences. Journal of Virology 78, 7465–7477.

Cohen S, Antignus Y. 1994. Tomato yellow leaf curl virus, a whitefly-borne geminivirus of tomatoes. Adv. Dis. Vector Res. 10, 259–88.

Cohen S, Regev A, Lavi S. 1997. Small polydispersed circular DNA (spcDNA) in human cells: association with genomic instability. Oncogene 14, 977-985.

Cohen S, Yacobi K, Segal D. 2003. Extrachromosomal circular DNA of tandemly repeated genomic sequences in Drosophila. Genome Research 13, 1133-1145.

Coursey T, Regedanz E, Bisaro DM. 2018. Arabidopsis RNA Polymerase V mediates enhanced compaction and silencing of geminivirus and transposon chromatin during host recovery from infection. Journal of Virology 92.

Dean FB, Nelson JR, Giesler TL, Lasken RS. 2001. Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. Genome Research 11, 1095-1099.

Dellaporta SL, Wood J, Hicks JB. 1983. A plant DNA minipreparation, Version II. Plant Molecular Biology Reporter. Volume 1, Issue 4, 19-21.

Deuschle K, Kepp G, Jeske H. 2016. Differential methylation of the circular DNA in geminiviral minichromosomes. Virology 499, 243–258.

Díaz-Pendón JA, Li F, Li WX, Ding, SW. 2007. Suppression of antiviral silencing by cucumber mosaic virus 2b protein in Arabidopsis is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. Plant Cell 19, 2053-63.

Díaz-Pendón JA, Cañizares MC, Moriones E, Bejarano ER, Czosnek H, Navas-Castillo J. 2010. Tomato yellow leaf curl viruses, ménage à trois between the virus complex, the plant and the whitefly vector. Molecular Plant Pathology 11, 441–450.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR, ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Donaire L, Wang Y, Gonzalez-Ibeas D, Mayer KF, Aranda MA, Llave C. 2009. Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. Virology 392, 203–214.

Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, Vashisht AA, Terragni J, Chin HG, Tu A, et al. 2012. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. Cell 151, 167–180.

Dvorackova M, Fojtova M, Fajkus J. 2015. Chromatin dynamics of plant telomeres and ribosomal genes. Plant Journal 83,18 –37.

Egelkrout EM, Robertson D and Hanley-Bowdoin L. 2001. Proliferating cell nuclear antigen transcription is repressed through an E2F consensus element and activated by geminivirus infection in mature leaves. The Plant Cell 13, 1437-1452.

Elmer JS, Sunter G, Gardiner WE, Brand L, Browning CK, Bisaro DM, Rogers SG. 1988. Agrobacterium-mediated inoculation of plants with Tomato golden mosaic virus DNAs. Plant Molecular Biology 10, 1-10.

Ermak G, Paszkowski U, Wohlmuth M, Mittelsten Scheid O, Paszkowski J. 1993. Cytosine methylation inhibits replication of African cassava mosaic virus by two distinct mechanisms. Nucleic Acids Research 21, 3445–3450.

Fernández-Calvino L, Osorio S, Hernández ML, Hamada IB, Del Toro FJ, Donaire L, Yu A, Bustos R, Fernie AR, Martínez-Rivas JM and Llave C. 2014. Virus-Induced Alterations in Primary Metabolism Modulate Susceptibility to Tobacco rattle virus in Arabidopsis. Plant Physiology 166,1821-1838.

Fondong VN. 2013. Geminivirus protein structure and function. Molecular Plant Pathology 14, 635–649.

Fransz P, Armstrong S, Alonso-Blanco C, Fischer TC, Torres-Ruiz RA, Jones G. 1998. Cytogenetics for the model system Arabidopsis thaliana. Plant J. 13, 867–876.

García-Andrés S, Monci F, Navas-Castillo J, Moriones E. 2006. Begomovirus genetic diversity in the native plant reservoir Solanum nigrum: evidence for the presence of a new virus species of recombinant nature. Virology 350, 433–42.

Gagnon-Kugler T, Langlois F, Stefanovsky V, Lessard F, Moss T. 2009. Loss of human ribosomal gene CpG methylation enhances cryptic RNA polymerase II transcription and disrupts ribosomal RNA processing. Mol Cell 35,414 – 425.

Gaubatz. 1990. Extrachromosomal circular DNAs and genomic sequence plasticity in eukaryiotic cells. Mutat Res. 237, 271-292.

Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL. 2006. DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. Cell 124, 495–506.

Gehring M. 2013. Genomic imprinting: insights from plants. Annual Reviews of Genetics 47, 187-208.

Ghoshal B, Sanfaçon H. 2015. Symptom recovery in virus-infected plants, revisiting the role of RNA silencing mechanisms. Virology 479-480, 167–179.

Gong Z, Morales-Ruíz T, Ariza RR, Roldán-Arjona T, David L, Zhu JK. 2002. ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell 111, 803-814.

Grossniklaus U, Kelly WG, Kelly B, Ferguson-Smith AC, Pembrey M, Lindquist S. 2013. Transgenerational epigenetic inheritance: how important is it? Nat Rev Genet. 14, 228-235.

Grummt I, Pikaard CS. 2003. Epigenetic silencing of RNA polymerase I transcription. Nature Reviews Molecular Cell Biology 4, 641-649.

Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D. 1999. Geminiviruses, models for plant DNA replication, transcription, and cell cycle regulation. Critical reviews in biochemistry and molecular biology 35, 105–140.

Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S. 2013. Geminiviruses, masters at

redirecting and reprogramming plant processes. Nature Reviews Microbiology 11, 777–788.

Henras AK, Soudet J, Gérus M, Lebaron S, Caizergues-Ferrer M, Mougin A, Henry Y. 2008. Cell Mol Life Sci 65, 2334-59.

Henras AK, Plisson-Chastang C, ODonohue M, Chakraborty A, Gleizes PE. 2015. An overview of pre-ribosomal RNA processing in eukaryotes. Wires RNA. doi. org/10.1002/wrna.1269.

Hirt B. 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26, 365–369.

Huang J, Ma L, Yang F, Fei SZ, Li L. 2008. 45S rDNA regions are chromosome fragile sites expressed as gaps in vitro on metaphase chromosomes of root-tip meristematic cells in Lolium spp. PLoS ONE, 3, e2167.

Iskandar HM, Simpson RS, Casu RE, Bonnett GD, Maclean DJ, Manners JM. 2004. Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane. Plant Molecular Biology Reporter 22, 325-337.

Jackel JN, Storer JM, Coursey T, Bisaro DM. 2016. Arabidopsis RNA polymerases IV and V are required to establish H3K9 methylation, but not cytosine methylation, on geminivirus chromatin. Journal of Virology 90, 7529–7540.

Jeske H. 2009. Geminiviruses. Current topics in microbiology and immunology 331, 185–226.

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE. 2007. The SRA methyl-cytosine-binding domain links DNA and histone methylation. Current Biology 17, 379-84.

Kashina BD, Mabagala RB, Mpunami A. 2003. Biomolecular relationships among isolates of Tomato yellow leaf curl tanzania virus. Phytoparasitica 31, 188-199.

Kaul S, Koo HL, Jenkins J, Rizzo M, Rooney T, Tallon LJ et al. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796-815.

Kim VN. 2008. Sorting out small RNAs. Cell 133, 25–26.

Kim J, Kil EJ, Kim S, Seo H, Byun HS, Park J, Chung MN, Kwak HR, Kim MK, Kim CS, Wang JW, Lee KY, Choi HS, Lee S. 2015. Seed transmission of Sweet potato leaf curl virus in sweet potato (Ipomoea batatas). Plant Pathology, doi, 10.1111/ppa.12366.

Kil EJ, Kim S, Lee YJ, Buyn HS, Park J, Seo H, Kim CS, Shim JK, Lee JH, Kim JK, Lee KY, Choi HS, Lee S. 2016. Tomato yellow leaf curl virus (TYLCV-IL), a seed-transmissible geminivirus in tomatoes. Scientific Reports 6, 19013.

Kobayashi T. 2008. A new role of the rDNA and nucleolus in the nucleus-rDNA instability maintains genome integrity. Bioessays 30, 267-272.

Koukalová B, Reich J, Matyásek R, Kuhrová V, Bezdek M. 1989. A BamHI family of highly repeated DNA sequences of Nicotiana tabacum. Theor Appl Genet 78, 77-80.

Kovarik A, Koukalová B, Lim KY, Matyasek R, Lichtenstein CP, Leitch AR, Bezdek M. 2000. Comparative analysis of DNA methylation in tobacco heterochromatic sequences. Chromosome Research 8, 527-541.

Kovarik A, Dadejova M, Lim YK, Chase MW, Clarkson JJ, Knapp S, Leitch AR. 2008. Evolution of rDNA in Nicotiana allopolyploids: a potential link between rDNA homogenization and epigenetics. Annals of Botany 101, 815-823.

Kravchik M, Damodharan S, Stav R, Arazi T. 2014. Generation and characterization of a tomato DCL3-silencing mutant. Molecular plant-microbe interactions, 221-222, 81–89.

Krueger F, Andrews SR. 2011. Bismark, a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571–1572.

Kumagai MH, Donson J, DellaCioppa G, Harvey D , Hanley K , Grill LK. 1995. Cytoplasmic inhibition of carotenoid biosynthesis with virus derived RNA. Proc. Natl. Acad. Sci. USA 92, 1679-1683.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Software, 1–10.

Law JA, Jacobsen SE. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature Reviews Genetics 11, 204–220.

Li B, Dewey CN. 2011. RSEM, accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. 2008. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133,523–536.

Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. 1998. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nature Genetics 19, 225-232.

Lozano-Durán R, Rosas-Diaz T, Luna AP, Bejarano ER, 2011. Identification of host genes involved in geminivirus infection using a reverse genetics approach. PLoS One 6, e22383.

Lukowitz W, Gilmor CS, Scheible WR. 2000. Positional cloning in Arabidopsis. Why it feels good to have a genome initiative working for you. Plant Physiology 123, 795-805.

Mallory A, Vaucheret H. 2010. Form, function, and regulation of Argonaute proteins. The Plant Cell 22, 3879–3889.

Maloy SR. 1990. Experimental techniques in bacterial genetics. Gones and Bartlett, Boston.

Mansoor S, Briddon RW, Zafar Y, Stanley J. 2003. Geminivirus disease complexes, an emerging threat. Trends in plant science 8, 128-134.

Matzke MA, Mosher RA. 2014. RNA-directed DNA methylation, an epigenetic pathway of increasing complexity. Nature Reviews Genetics 15, 394–408.

Martínez G, Castellano M, Tortosa M, Pallás V, Gómez G. 2014. A pathogenic non-coding RNA induces changes in dynamic DNA methylation of ribosomal RNA genes in host plants. Nucleic Acid Research 42, 1553-62.

Mason G, Caciagli P, Accotto GP, Noris E. 2007. Real-time PCR for the quantitation of Tomato yellow leaf curl Sardinia virus in tomato plants and in Bemisia tabaci. Journal of Virological Methods 147, 282-289.

Mathieu O, Yukawa Y, Sugiura M, Picard G, Tourmente S. 2002. 5S rRNA gene expression is not inhibited by DNA methylation in Arabidopsis. Plant Journal 29, 313–323.

Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H et al. 2008. Sorting of small RNAs into Arabidopsis Argonaute complexes is directed by the 5' terminal nucleotide. Cell 133, 116–127.

Miozzi L, Pantaleo V, Burgyán J, Accotto GP, Noris E. 2013. Analysis of small RNAs derived from Tomato yellow leaf curl Sardinia virus reveals a cross reaction between the major viral hotspot and the plant host genome. Virus Research 178, 287–296.

Moissiard G, Voinnet O. 2006. RNA silencing of host transcripts by cauliflower mosaic virus requires coordinated action of the four Arabidopsis Dicer-like proteins. Proc Natl Acad Sci USA 103, 112-134.

Monci F, Sánchez-Campos S, Navas-Castillo J, Moriones E. 2002. A natural recombinant between the geminiviruses Tomato yellow leaf curl Sardinia virus and Tomato yellow leaf curl virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. Virology 303, 317–26.

Morales-Ruiz T, Ortega-Galisteo AP, Ponferrada-Marin MI, Martinez-Macias MI, Ariza RR, Roldan-Arjona T. 2006. DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. Proc Natl Acad Sci USA 103, 6853–6858.

Morilla G, Janssen D, García-Andrés S, Moriones E, Cuadrado IM, Bejarano ER. 2005. Pepper is a dead-end host for Tomato yellow leaf curl virus. Phytopathology 95, 1089–1097.

Morilla G, Castillo AG, Preiss W, Jeske H, Bejarano ER, 2006. A versatile transreplication-based system to identify cellular proteins involved in geminivirus replication. Journal of Virology 80, 3624-3633.

Moriones E, Arno J, Accotto GP, Noris E, Cavallarin L. 1993. First report of tomato yellow leaf curl virus in Spain. Plant Disease. 77, 953.

Murata M, Heslop-Harrison JS, Motoyoshi F. 1997. Physical mapping of the 5S ribosomal RNA genes in Arabidopsis thaliana by multi-color fluorescence in situ hybridization with cosmid clones. Plant Journal. 12, 31–37.

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. The Plant Cell 7, 705-719.

Navas-Castillo J, Sánchez-Campos S, Díaz JA, Sáez-Alonso E, Moriones E. 1997. First report of Tomato yellow leaf curl virus-Is in Spain: Coexistence of two different geminiviruses in the same epidemic outbreak. Plant disease 81, 1461.

Navas-Castillo J, Sánchez-Campos S, Díaz JA, Sáez-Alonso E, Moriones E. 1999. Tomato Yellow Leaf Curl Virus-Is Causes a Novel Disease of Common Bean and Severe Epidemics in Tomato in Spain. Plant Disease 83, 29-32.

Navas-Castillo J, Sánchez-Campos S, Noris E, Louro D, Accotto GP, Moriones E. 2000. Natural recombination between Tomato yellow leaf curl virus-Is and Tomato leaf curl virus. Journal of General Virology 81, 2797-801.

Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S. 2011. Emerging virus diseases transmitted by whiteflies. Annu Rev Phytopathol 49, 219-248.

Noris E, Lucioli A, Tavazza R, Caciagli P, Accotto GP, Tavazza M. 2004. Tomato yellow leaf curl Sardinia virus can overcome transgene-mediated RNA silencing of two essential viral genes. Journal of General Virology 85, 1745–1749.

Ortega-Galisteo AP, Molares-Ruíz T, Ariza RR, Roldán-Arjona T. 2008. Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. Plant Mol Biol 67, 671-81.

Paprotka T, Deuschle K, Metzler V, Jeske H. 2011. Conformation-selective methylation of Geminivirus DNA. Journal of Virology 85, 12001–12012.

Penterman J, Uzawa R, Fischer RL. 2007. Genetic interactions between DNA demethylation and methylation in Arabidopsis. Plant Physiol 145, 1549–1557.

Pereira-Carvalho R, Díaz-Pendón J, Fonseca M, Boiteux L, Fernández-Muñoz R, Moriones E,

Resende R. 2015. Recessive resistance derived from Tomato cv. Tyking limits drastically the spread of Tomato yellow leaf curl virus. Viruses 7, 2518–2533.

Pierce EJ and Rey ME. 2013. Assessing global transcriptome changes in response to South African cassava mosaic virus infection in susceptible Arabidopsis thaliana. Plos One 8, e67534.

Pikaard CS, Mittelsten O. 2014. Epigenetic regulation in plants. Cold Spring Harbor Perspectives in Biology. doi, 10.1101/cshperspect.a019315.

Pilartz M, Jeske H. 1992. Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. Virology 189, 800–802.

Pilartz M, Jeske H. 2003. Mapping of abutilon mosaic geminivirus minichromosomes. Journal of Virology 77, 10808–10818.

Pontvianne F, Blevins T, Chandrasekhara C, Feng W, Stroud H, Jacobsen SE, Michaels SD, Pikaard CS. 2012. Histone methyltransferases regulating rRNA gene dose and dosage control in Arabidopsis. Genes Dev. 26, 945–957.

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

Pradeep. 1993. Extraction of episomal DNA from procyclics. EMBO Journal 12, 2529-2533.

Preiss W, Jeske H. 2003. Multitasking in replication is common among geminiviruses. Journal of Virology 77, 2972–2980.

Pumplin N, Voinnet O. 2013. RNA silencing suppression by plant pathogens, defence, counter-defence and counter-counter-defence. Nature Reviews Microbiology 11, 745–760.

Quinlan AR, Hall IM. 2010. BEDTools, a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

Raja P, Jackel JN, Li S, Heard IM, Bisaro DM. 2014. Arabidopsis double-stranded RNA binding protein DRB3 participates in methylation-mediated defense against geminiviruses. Journal of Virology 88, 2611–2622.

Raja P, Sanville BC, Buchmann RC, Bisaro DM. 2008. Viral genome methylation as an epigenetic defense against geminiviruses. Journal of Virology 82, 8997–9007.

Raja P, Wolf JN, Bisaro DM. 2010. RNA silencing directed against geminiviruses, post-transcriptional and epigenetic components. Biochimica et biophysica acta 1799, 337–351.

Rajabu CA, Kennedy GG, Ndunguru J, Ateka EM, Tairo F, Hanley-Bowdoin L, Ascencio-Ibáñez JT. 2018. Lanai: a small, fast growing tomato variety is an excellent model system for studying

geminiviruses. Journal of Virological Methods 256, 89-99.

Ramesh SV, Sahu PP, Prasad M, Praveen S, Pappu HR. 2017. Geminiviruses and plant hosts, a closer examination of the molecular arms race. Viruses 9.

Ratcliff F, Martin-Hernandez AM, Baulcombe DC. 2001. Technical Advance. Tobacco rattle virus as a vector for analysis of gene function by silencing. The Plant Journal for cell and molecular biology 25, 237-245.

Reina J, Morilla G, Rodríguez MD, Jansen J, Cuadrado IM, Bejarano ER. 2000. First report of Capsicum annun plants infected by Tomato yellow leaf curl virus. Plant Disease 83, 1176.

Rizvi, I., Choudhury, N.R., Tuteja, N. 2014. Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication. Archives of Virology. doi 10.1007/s00705-014-2297-7.

Rodriguez-Negrete EA, Carrillo-Tripp J, Rivera-Bustamante RF. 2009. RNA silencing against geminivirus, complementary action of posttranscriptional gene silencing and transcriptional gene silencing in host recovery. Journal of Virology 83, 1332–1340.

Rodríguez-Negrete EA, Lozano-Durán R, Piedra-Aguilera A, Cruzado L, Bejarano ER, Castillo AG. 2013. Geminivirus Rep protein interferes with the plant DNA methylation machinery and suppresses transcriptional gene silencing. New Phytologist 199, 464–475.

Rodríguez-Negrete EA, Sánchez-Campos S, Cañizares MC, Navas-Castillo J, Moriones E, Bejarano ER, Grande-Pérez A. 2014. A sensitive method for the quantification of virion-sense and complementary-sense DNA strands of circular single-stranded DNA viruses. Scientific Reports 4, 1851–1859.

Rogans SJ, Allie F, Tirant JE, Rey MEC. 2016. Small RNA and methylation responses in susceptible and tolerant landraces of cassava infected with South African cassava mosaic virus. Virus Research 225, 10–22.

Rojas MR, Jiang H, Salati R, Xoconostle-Cazares B, Sudarshana MR, Lucas WJ, Gilbertson RL. 2001. Functional analysis of proteins involved in movement of the monopartite begomovirus, Tomato yellow leaf curl virus. Virology 291, 110-125.

Rojas MR, Hagen C, Lucas WJ, Gilbertson RL. 2005. Exploiting chinks in the plant's armor, evolution and emergence of geminiviruses. Annual review of phytopathology 43, 361- 394.

Rosas-Díaz T, Zhang D, Lozano-Durán R. 2017. No evidence of seed transmissibility of Tomato yellow leaf curl virus in Nicotiana benthamiana. J Zheijiang Univ-Sci B 18, 437-440.

Rotenberg D, Thompson TS, German TL, Willis DK. 2006. Methods for effective real-time rt-pcr analysis of virus-induced gene silencing. Journal of Virological Methods 138, 49-59.

Ruíz MT, Voinnet O, Baulcombe DC. 1998. Initiation and maintenance of virus-induced gene silencing. Plant Cell, 10, 937-946.

Sahu PP, Sharma N, Puranik S, Prasad M. 2014. Post-transcriptional and epigenetic arms of RNA silencing, a defense machinery of naturally tolerant tomato plant against Tomato leaf curl New Delhi virus. Plant Molecular Biology Reporter 32, 1015–1029.

Sambrook J, Russell DW. 2001. Molecular cloning, a laboratory manual, 3rd edn. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.

Sánchez-Durán MA, Dallas MB, Ascencio-Ibáñez JT, Reyes mI, Arroyo-Mateos M, Ruíz-Albert J, Hanely-Bowdoin L, Bejarano ER 2011. Interaction between geminivirus replication protein and the SUMO-conjugating enzyme is required for viral infection. Journal of Virology 85, 9789-800.

Saze H, Mittelsten O, Paszkowski J. 2003. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nature Genetics 34, 65-69.

Seguin J, Otten P, Baerlocher L, Farinelli L, Pooggin MM. 2016. MISIS-2, A bioinformatics tool for in-depth analysis of small RNAs and representation of consensus master genome in viral quasispecies. Journal of Virological Methods 233, 37–40.

Seo JK, Kim MK, Kwak HR, Choi HS, Nam M, Choe J et al. 2018. Molecular dissection of distinct symptoms induced by Tomato chlorosis virus and and Tomato yellow leaf curl virus based on comparative transcriptome analysis. Virology 516, 1-20.

Shaw P, Mckeown P. 2011. The structure of rDNA chromatin. In Olson MOJ (Ed). The nucleolus. Springer New York, 43-55.

Shivaprasad PV, Akbergenov R, Trinks D, Rajeswaran R, Veluthambi K, Hohn T, Pooggin MM. 2005. Promoters, transcripts, and regulatory proteins of Mungbean yellow mosaic geminivirus. Journal of Virology 79, 8149–8163.

Srivastava R, Srivastava R, Ahn SH. 2016. The epigenetic pathways to ribosomal DNA silencing. Microbiology and Molecular Biology Reviews 80.

Stanley J, Markham PG, Callis RJ, Pinner MS 1986. The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. EMBO Journal 5, 1761-1767.

Stroud H, Do T, Du J, Zhong X, Feng S, Johnson L, Patel DJ, Jacobsen SE. 2014. Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nature Structural and Molecular Biology 21, 64–72.

Stroud H, Greenberg MVC, Feng S, Bernatavichute YV, Jacobsen SE. 2013. Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell 152, 352–364.

Sun YW, Tee CS, Ma YH, Wang G, Yao, XM, Ye J. 2015. Attenuation of histone methyltransferase Kryptonite-mediated transcriptional gene silencing by geminivirus. Scientific Reports. doi, 10.1038/srep16476.

Takeuchi Y, Horiuchi T, Kobayashi T. 2003. Transcription-dependent recombination and the role of fork collision in yeast rDNA. Genes Dev 17,1497 – 1506.

Torchetti EM, Pegoraro M, Navarro B, Catoni M, Di Serio F, Noris E. 2016. A nuclear-replicating viroid antagonizes infectivity and accumulation of a geminivirus by upregulating methylation-related genes and inducing hypermethylation of viral DNA. Scientific Reports, 1–16.

Tu J, Sunter G. 2007. A conserved binding site within the Tomato golden mosaic virus AL-1629 promoter is necessary for expression of viral genes important for pathogenesis. Virology 367, 117–125.

Tucker S, Vitins A, Pikaard CS. 2010. Nucleolar dominance and ribosomal RNA gene silencing. Curr. Opin. Cell Biol. 22, 351–356.

Wang H, Hao L, Shung C, Sunter G, Bisaro DM. 2003. Adenosine kinase is inactivated by geminivirus AL2 and L2 proteins. Plant Cell 15, 3020–3032.

Wang B, Li F, Huang C, Yang X, Qian Y, Xie Y, Zhou X. 2014. V2 of tomato yellow leaf curl virus can suppress methylation-mediated transcriptional gene silencing in plants. Journal of General Virology 95, 225-230.

Xu C, Sun X, Taylor A, Jiao C, Xu Y, Cai X, Wang X, Ge C, Pan G, Wang Q, et al. 2017. Diversity, distribution, and evolution of tomato viruses in China uncovered by small RNA sequencing. Journal of Virology 91, e00173–17.

Yadav RK, Chattopadhyay D. 2011. Enhanced viral intergenic region-specific short interfering RNA accumulation and DNA methylation correlates with resistance against a geminivirus. Molecular Plant-Microbe Interactions 24, 1189–1197.

Yang X, Wang Y, Guo W, Xie Y, Xie Q, Fan L, Zhou X. 2011a. Characterization of small interfering RNAs derived from the geminivirus/betasatellite complex using deep sequencing. PLoS ONE 6, e16928.

Yang X, Xie Y, Raja P, Li S, Wolf JN, Shen Q, Bisaro DM, Zhou X. 2011b. Suppression of methylation-mediated transcriptional gene silencing by βC1-SAHH protein interaction during geminivirus-betasatellite infection. PLoS Pathogens 7, e1002329.

Yang LP, Fang YY, An CP, Dong L, Zhang ZH, Chen H, Xie Q, Guo HS. 2013. C2-mediated decrease in DNA methylation, accumulation of siRNAs, and increase in expression for genes involved in defense pathways in plants infected with Beet severe curly top virus. The Plant J 73, 910-917.

Yusupova G, Yusupov M. 2014. High-resolution structure of the eukaryotic 80S ribosome. Annu Rev Biochem 83, 467–486.

Zemach A, Kim MY, Hsieh P-H, Coleman-Derr D, Eshed-Williams L, Thao K, Harmer SL, Zilberman D. 2013. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell 153, 193–205.

Zerbini FM, Briddon RW, Idris A, Martin DP, Moriones E, Navas-Castillo J, Rivera-Bustamante R, Roumagnac P, Varsani A, Report I. 2017. ICTV Virus Taxonomy Profile, Geminiviridae. Journal of General Virology 98, 131–133.

Zhang W, Olson N, Baker T, Faulkner L, Agbandje-McKenna M, Boulton M, Davies, J, McKenna R. 2001. Structure of the Maize Streak Virus gemininate particle. Virology 279, 471-477.

Zhang Z, Chen H, Huang X, Xia R, Zhao Q, Lai J, Teng K, Li Y, Liang L, Du Q, et al. 2011. BSCTV C2 attenuates the degradation of SAMDC1 to suppress DNA methylation-mediated gene silencing in Arabidopsis. The Plant Cell 23, 273–288.

Zhou XP. 2013. Advances in understanding begomovirus satellites. Annu Rev Phyt 51, 357–381.

Resumen en español.

Introducción general.

Los geminivirus constituyen una amplia familia de virus de plantas transmitidos por insectos, contienen una (monopartitos) o dos (bipartitos) moléculas de DNA de cadena sencilla (*single stranded* DNA, ssDNA) que son empaquetadas en dos partículas gemelas fusionadas que forman la cápside. El geminivirus es uno de los patógenos de plantas más devastadores en el mundo, causando importantes pérdidas de cosechas en plantas de tomate, maíz, algodón, legumbres, cassava, etc. Síntomas comunes asociados con la infección geminiviral incluyen menor tamaño de las plantas, malformación de los órganos reproductivos o curvatura y amarilleamiento de las hojas. Estos virus pertenecen a la familia *Geminiviridae*, la cual puede clasificarse en siete géneros de acuerdo al insecto vector y a la organización del genoma, siendo el género *Begomovirus* el principal objeto de estudio de esta Tesis Doctoral. El genoma de los geminivirus codifica seis a ocho solapantes marcos abiertos de lectura (*open reading frames*, ORFs) que son transcritos bidireccionalmente a partir de promotores localizados en la región intergénica (*intergenic region*, IR).

Los begomovirus infectan plantas dicotiledóneas, son transmitidos por la mosca blanca *Bemisia tabaci*, y presentan genomas monopartitos o bipartitos. El genoma de los begomovirus monopartitos está organizado en dos unidades transcripcionales (*Left*, L and *Right*, R) separadas por la IR que contiene los elementos genéticos para el inicio de la replicación y la transcripción del genoma viral. La cadena virión (*virion strand*, VS) contiene dos ORFs expresados desde la unidad transcripcional R: V2 y V1, mientras que la cadena complementaria (*complementary strand*, CS) contiene cuatro ORFs: C1, C2, C3 y C4 que son expresados desde la unidad transcripcional L. C1 codifica la proteína iniciadora de la replicación, Rep, siendo la única esencial para la replicación del geminivirus. Por otro lado, los begomovirus bipartitos contienen dos moléculas de DNA (DNA-A y DNA-B); el DNA-A es homólogo al genoma de los monopartitos, y el DNA-B contiene dos ORFs adicionales: BV1 y BC1.

Durante el ciclo de vida de los geminivirus, la planta adquiere los viriones a través del insecto vector; posteriormente, el ssDNA viral es liberado al núcleo de la planta, donde es convertido en DNA de doble cadena (*double stranded* DNA, dsDNA) para formar minicromosomas tras el ensamblaje con histonas de la planta. Dicha molécula de dsDNA puede ser replicada tras la síntesis de Rep, mediante dos principales mecanismos: replicación por círculo rodante (*rolling-circle replication*, RCR) y replicación dependiente de recombinación (*recombination-dependent replication*, RDR). Las moléculas resultantes pueden alcanzar otras células de la planta o pasar nuevamente al insecto vector.

Durante la infección, los geminivirus interfieren con la maquinaria celular de la planta mediante múltiples interacciones a través de las proteínas que codifica. Estos virus no codifican

su propia maquinaria de replicación, e infectan células diferenciadas que no se dividen, por lo que inducen la reprogramación del ciclo celular de la planta para mediar su propia replicación. Por otro lado, los geminivirus deben confrontar mecanismos de defensa tales como el silenciamiento génico.

El silenciamiento génico mediado por RNA (RNA *silencing*) juega un importante papel en el desarrollo de la planta y la defensa ante estreses bióticos y abióticos. Es desencadenado por moléculas de RNA de doble cadena (*double stranded* RNA, dsRNA), las cuales son procesadas por ribonucleasas llamadas *Dicer-like* (DCL) y convertidas en moléculas de dsRNA de 20-25 nucleótidos (nt) llamadas RNAs pequeños (*small* RNAs, sRNAs). Una de las cadenas de dichas moléculas es cargada en un complejo proteico que incluye Argonauta (AGO). En el citoplasma, cuando el complejo AGO-sRNA encuentra complementariedad de bases con un RNA mensajero (*messenger* RNA, mRNA), desencadena su degradación o inhibición de la traducción, induciendo así su silenciamiento a nivel post-trancripcional (*post-transcriptional gene silencing*, PTGS). En el núcleo, este complejo puede servir como guía para deposición de marcas represivas en la cromatina, como la metilación del DNA o modificaciones de las histonas, en aquellas regiones del DNA que son complementarias al sRNA, induciendo su silenciamiento a nivel transcripcional (*transcriptional gene silencing*, TGS). En plantas, la metilación del DNA ocurre en las citosinas, pudiendo darse en tres contextos de secuencia: CG, CHG y CHH (H: C, T o A) y en todos ellos se promueve un estado compactado y silenciado de la cromatina. La metilación del DNA mediada por RNA (*RNA-directed DNA methylation*, RdDM) induce la adición de grupos metilo en nuevas secuencias de DNA que previamente carecían de dicha marca epigenética. En este mecanismo participan, entre otras proteínas, dos RNA polimerasas dependientes de DNA, Pol IV y Pol V, sRNAs de 24-nt, y dos metiltransferasas, DRM2 (DOMAIN REARRANGED METHYLTRANSFERASE 2) y DRM1. Una vez se ha establecido la metilación en cierta secuencia, ésta debe ser mantenida tras las sucesivas rondas de replicación del DNA. Esto es llevado a cabo a través de metiltransferasas de mantenimiento: MET1 (METHYLTRANSFERASE 1, metilación en CG), CMT3 (CHROMOMETHYLASE 3, metilación en CHG), CMT2 (metilación en CHH) y DRM2 (metilación en CHH). Por otro lado, la metilación del DNA puede ser eliminada de forma pasiva mediante la ausencia de metilación tras la replicación, o de forma activa mediante DNA glicosilasas tales como ROS1 (REPRESSOR OF SILENCING).

El silenciamiento génico constituye un mecanismo clave en la defensa antiviral de la planta. Moléculas de dsRNA pueden generarse a partir del genoma de virus de RNA o de DNA, y por tanto producir sRNAs derivados de virus, llamados vsRNAs (*viral small* RNAs). Los vsRNAs pueden ser cargados en proteínas AGO y dirigir la degradación de RNAs virales o la metilación del DNA viral. El DNA geminiviral es metilado durante la infección, siendo el grado de metilación variable (50% a 2%) y dependiente del sistema geminivirus-planta en cuestión. En plantas infectadas por geminivirus se ha descrito la presencia de vsRNAs de 20 a 25-nt, lo cual sugiere que la metilación del DNA viral puede ser inducida mediante RdDM. Adicionalmente, ya que el genoma del geminivirus se asocia con histonas de la planta formando un minicromosoma, también dichas proteínas contienen marcas post-traduccionales que regulan la cromatina viral.

Numerosas observaciones han sugerido que la metilación del DNA conforma un mecanismo de defensa de la planta ante estos patógenos. Por un lado, se ha observado que tejidos asintomáticos emergentes en una planta infectada contienen DNA viral hipermetilado; por otro lado, plantas deficientes en la maquinaria de metilación del DNA muestran mayor susceptibilidad a la infección. Otros trabajos han descrito que al menos tres proteínas codificadas en el genoma geminiviral (C2, Rep y V2), y otras moléculas de DNA llamadas beta-satélites que se asocian con el geminivirus, funcionan como supresores de silenciamiento génico transcripcional, lo que nuevamente sugiere una función antiviral de la metilación del DNA.

Objetivos.

1. Generación de los mapas transcriptómicos y epigenéticos del geminivirus TYLCV en tomate.

2. Determinar la relevancia biológica de la metilación del DNA durante una infección por geminivirus en *Arabidopsis thaliana* y *Nicotiana benthamiana*.

3. Analizar la transmisión del geminivirus a la descendencia a través de semilla en *Arabidopsis thaliana*.

Estos tres objetivos se han desarrollado en los cuatro capítulos que se resumen a continuación.

Capítulo 1. Mapas integrados con resolución nucleotídica del transcriptoma, sRNAoma y metiloma de TYLCV en tomate.

Con el objetivo de describir en un mismo estudio la relación existente entre la expresión génica, el patrón de metilación del genoma y los sRNAs derivados (vsRNAs) del geminivirus TYLCV (*Tomato yellow leaf curl virus*), obtuvimos el transcriptoma, sRNAoma y metiloma de dicho begomovirus durante una infección en plantas de tomate. Para ello se usó la tecnología de secuenciación Illumina (RNA-*sequencing* (RNA-Seq) y *small* RNA-*sequencing* (sRNA-Seq)), y tratamiento con bisulfito más secuenciación (BS-Seq). Dicho análisis fue llevado a cabo a diferentes tiempos de infección y con dos métodos de inoculación del virus: mediante el uso del vector natural, la mosca blanca *Bemisia tabaci*, y mediante el uso de la bacteria *Agrobacterium tumefaciens*, método de infección ampliamente usado. Esta aproximación nos ha permitido obtener de forma integrada el mapa epigenético y transcriptómico de TYLCV, a resolución de nucleótido.

Plantas de tomate adultas fueron inoculadas con TYLCV mediante *Agrobacterium* o *Bemisia*. Como controles negativos, grupos diferentes de plantas fueron inoculadas con moscas o bacterias que no portaban virus. El tejido apical de dichas plantas fue colectado a 2, 7, 14 y 21 días post-inoculación (dpi), y posteriormente procesado para extracción de DNA y RNA total. El patrón de acumulación de DNA viral fue similar en plantas tratadas con *Agrobacterium* o *Bemisia*, habiéndose observado un incremento exponencial desde 7 dpi hasta 14 dpi, donde se

alcanzó el máximo nivel de DNA viral hasta 21 dpi. Dicha acumulación fue caracterizada en profundidad mediante la cuantificación de ambas cadenas del geminivirus (VS y CS). Acorde con trabajos previos, VS alcanzó niveles más altos que CS (13 a 22 veces más) en todos los tiempos durante la infección. En cuanto al análisis del mRNA viral, su cinética fue consistente con la acumulación de DNA, dándose el mayor incremento desde 7 dpi a 14 dpi, y manteniéndose hasta 21 dpi. Nuevamente no hubo diferencias significativas al comparar ambos métodos de inoculación.

Con el objetivo de determinar la distribución de transcritos virales acumulados durante la infección, las lecturas provenientes del RNA-Seq fueron mapeadas contra el genoma viral. Nuestros datos indicaron que estas lecturas correspondientes al mRNA viral no fueron distribuidas homogéneamente a lo largo del genoma, siendo casi inexistentes en la IR o en las cadenas antisentido de las unidades transcripcionales L y R. Los transcritos de la cadena VS se acumularon a mayores niveles que los provenientes de la cadena CS; las lecturas correspondientes a V1 y V2 representaron el 75% del total en cualquier tiempo de la infección.

La secuenciación de sRNAs determinó que el número total de vsRNAs de 20 a 25-nt aumentó a lo largo de la infección, siguiendo una cinética acorde a lo observado para la acumulación de DNA y mRNA viral, y siendo similar para ambos métodos de inoculación. El nivel de redundancia de vsRNAs también aumentó a lo largo de la infección, ya que el número de lecturas totales (redundantes) tuvo un mayor incremento que el número de lecturas únicas. Los tamaños de vsRNAs con mayor número de lecturas totales fueron 21 y 22-nt, representando el 80% del total de vsRNAs de 20-25-nt para cualquier tiempo analizado; el patrón de acumulación también fue mantenido para 20, 23, 24 y 25-nt. Por otro lado, los porcentajes de vsRNAs de 20 a 25-nt únicos no fueron mantenidos durante la infección; aunque los de 21 y 22-nt fueron los más abundantes en todos los tiempos, su proporción fue reducida desde el 60% a 7 dpi hasta el 40% a 21 dpi, con respecto al total de vsRNAs únicos de 20 a 25-nt. Nuestros datos indicaron que la mayoría de la diversidad de vsRNA ya existe a 7 dpi. El significativo incremento de vsRNAs observado entre 7 y 14 dpi es fundamentalmente debido a la producción de copias adicionales del mismo vsRNA, más que a la generación de nuevos vsRNAs únicos. Este aumento de la redundancia fue principalmente causado por el crecimiento del número total de vsRNAs de 21 y 22-nt durante la infección.

Para analizar la distribución genómica de los vsRNAs, las lecturas fueron mapeadas contra el genoma de TYLCV. Dicho análisis reveló que los vsRNAs de 20-25-nt se distribuyeron a lo largo de todo el genoma, con igual proporción en ambas cadenas (igual porcentaje de lecturas en sentido y antisentido) y patrón similar para ambos métodos de inoculación. Este resultado indica que los vsRNAs fueron generados a partir de precursores de dsRNA derivados de las dos cadenas. Tanto los vsRNAs en sentido como en antisentido tuvieron un patrón de distribución heterogéneo a lo largo del genoma del virus, con una mayor proporción concentrada en áreas específicas; dicha observación fue altamente consistente entre réplicas biológicas y tiempos analizados. Tres regiones correspondientes a las secuencias solapantes entre V2/V1, C1/C2/C3 y C1/C4 (OL1, OL2 y OL3 respectivamente) contuvieron más del 60% de los vsRNAs y mostraron la

densidad más alta por nucleótido. Trabajos previos también describieron la existencia de regiones del genoma del geminivirus, OL1 y OL3, con mayor número de vsRNAs, pero no en la región OL2. Estas regiones del genoma con mayor acumulación fueron alternadas con otras con menor abundancia, como la IR, la región limítrofe entre las unidades L y R, o secuencias promotoras y terminadoras. En cuanto a la composición de dichos puntos de mayor abundancia de vsRNAs, estuvieron representados por los tres tamaños mayoritarios, 21, 22 y 24-nt, lo cual sugiere que los mismos precursores de dsRNA pueden ser procesados por distintas DCLs. Sin embargo, un análisis más detallado reveló que, mientras que la distribución a lo largo del genoma viral de los vsRNAs de 21 y 22-nt fue muy similar, los vsRNAs de 24-nt mostraron una relativa mayor acumulación en la IR, durante toda la infección. El porcentaje de vsRNAs mapeados a la IR alcanzó el 9% para los de 24-nt, mientras que los de 21 y 22-nt mostraron hasta el 1.4% y 3.6% respectivamente.

El metiloma de TYLCV fue generado a través de Bs-Seq en muestras colectadas a 14 dpi. El porcentaje de citosinas metiladas alcanzó el 1% en plantas inoculadas con *Bemisia* y el 2,5% en plantas inoculadas con *Agrobacterium*, una diferencia que puede ser explicada por el ligero retraso en el desarrollo de la infección mediada por la mosca en comparación con la bacteria. A pesar de esta diferencia, la proporción relativa de citosinas metiladas en los tres contextos (CG, CHG y CHH) fue similar para ambos métodos, siendo el porcentaje de metilación en CHH y CG el más alto y el más bajo respectivamente. Ya que la principal diana de metilación del DNA geminiviral es el dsDNA, un valor más preciso del porcentaje de metilación sería estimado considerando el porcentaje de citosinas que forman parte de las moléculas de dsDNA viral. En base a nuestros datos sobre acumulación de VS y CS, el 11% y el 18% de las cadenas de DNA viral forman moléculas de dsDNA en plantas tratadas con *Agrobacterium* o *Bemisia*, respectivamente. Teniendo en cuenta estos resultados, estimamos que el porcentaje de citosinas metiladas en dsDNA viral está en torno al 23% en muestras tratadas con la bacteria, y el 7% en las tratadas con la mosca.

El análisis de la distribución de la metilación a lo largo del genoma de TYLCV nos permitió distinguir dos regiones con alta densidad de metilación, siendo altamente consistente entre réplicas biológicas y métodos de inoculación. Acorde con lo descrito para otros geminivirus, observamos altos niveles de metilación en la IR, los cuales se extendieron en ambos sentidos y cayeron drásticamente al comienzo de los ORFs correspondientes a V1 y C4. La alta resolución de nuestra aproximación nos permitió realizar un estudio más detallado de la metilación en dicha región. El nivel más bajo de metilación de la IR fue localizado en una secuencia que contiene el dominio de unión de C2; en cambio, el nivel más alto fue observado en la región comprendida entre el comienzo del ORF correspondiente a C1 y el comienzo del ORF de C4.

En conjunto, los resultados derivados de este capítulo nos permiten concluir que no existen diferencias significativas para ninguno de los parámetros analizados (acumulación de DNA o RNA viral, vsRNA o metilación de DNA viral) al comparar muestras infectadas mediante ambos métodos. El hecho de que elementos de la ruta RdDM parecen estar involucrados en la metilación del genoma geminiviral, junto con la existencia de vsRNAs de 24-nt observada en la

infección con TYLCV, sugiere que estos vsRNAs podrían estar dirigiendo la metilación del DNA del geminivirus. Con el objetivo de establecer conexiones entre los patrones de metilación, vsRNA y transcrito viral, estos datos fueron compilados y analizados de forma conjunta. Detectamos altos niveles de metilación en secuencias que contienen o son adyacentes a los tres promotores descritos en geminivirus, sin embargo, las secuencias promotoras y terminadoras mostraron baja abundancia de vsRNAs. En general, no hubo una clara correlación entre zonas de mayor acumulación de vsRNAs y picos de metilación del DNA viral. Tras analizar los patrones de metilación y distribución de vsRNAs pudimos apreciar tres distintivos escenarios epigenéticos. La región que contiene el promotor para la unidad transcripcional R mostró elevada metilación y alta densidad de vsRNAs de 24-nt, lo que sugiere la participación de la ruta RdDM en dicha región; la acumulación de vsRNAs de 21 y 22-nt fue baja. Un segundo escenario fue observado en el extremo 5' de C1, el cual mostró altos niveles de metilación, pero acumuló muy bajos niveles de vsRNAs de cualquier tamaño, sugiriendo que la metilación de esta región no ocurre a través de la ruta RdDM. Adicionalmente, detectamos un tercer perfil epigenético correspondiente al pico de metilación localizado en la región central de C1, donde altos niveles de metilación correlacionaron con elevada abundancia de vsRNAs de 24-nt y moderados niveles de 21 y 22-nt.

El tipo y cantidad de metilación del DNA y su localización dentro de un gen (promotor o ORF) determinan el impacto de esta marca epigenética en la transcripción. Los resultados obtenidos indicaron que en la población viral, ambos promotores bidireccionales dentro de la IR son metilados en distinto grado, lo que podría explicar las diferencias en expresión encontradas en las unidades transcripcionales R y L. V2 y V1 son expresados desde la misma unidad R, sin embargo nuestros datos de RNA-Seq mostraron una diferencial expresión de ambos; el hecho de que una amplia región en el extremo 5' de V2 mostró altos y densos niveles de metilación, sugiere que esta marca epigenética podría regular la menor acumulación de transcritos de V2 con respecto a V1. Igualmente podría regularse la expresión de los ORFs virales de la unidad L, cuya acumulación de transcrito no fue uniforme. Nuestro trabajo suscita nuevas preguntas acerca del rol de la metilación del DNA geminiviral como elemento regulador de su expresión.

Capítulo 2. Papel del la maquinaria de metilación del DNA de la planta durante una infección por geminivirus.

Trabajos previos han sugerido un rol defensivo antiviral de la metilación del DNA. En *A. thaliana*, mutantes simples de las metiltransferasas MET1 y CMT3 (involucradas en metilación de mantenimiento en CG y CHG, respectivamente), y el doble mutante en DRM1 y DRM2 (involucradas en metilación de mantenimiento en CHH y en metilación *de novo* mediada por RNA (RdDM) en los tres contextos de secuencia), mostraron una mayor susceptibilidad a la infección por geminivirus. En *N. benthamiana*, plantas en las que la expresión de *NbCMT3* se encontraba reprimida, mostraron mayor acumulación de geminivirus.

El objetivo de este capítulo es determinar la relevancia biológica de la maquinaria de metilación del DNA en la infección por el geminivirus TYLCV. Para ello infectamos mediante

agroinoculación plantas de *A. thaliana* deficientes en metilación del DNA (un mutante simple en *MET1* y un doble mutante en *DRM2* y *CMT3*) con TYLCV, y medimos mediante qPCR los niveles de acumulación viral. También se incluyeron en el análisis plantas mutantes en *ROS1*, gen que codifica una DNA glicosilasa que contribuye a la eliminación de la metilación del DNA. Los resultados obtenidos mostraron niveles similares de DNA viral al comparar las plantas mutantes y con las plantas silvestres infectadas. Durante la realización de dichos experimentos, nuestro grupo obtuvo nuevos datos sobre la infección por BCTV (*Beet curly top virus*) en *A. thaliana*, los cuales sugerían una mayor acumulación de DNA viral en tejido floral que en hojas de la roseta. Este resultado fue confirmado tras medir por qPCR los niveles de DNA de del curtovirus BCTV y adicionalmente TYLCV, en tejido floral y hojas apicales de la roseta de plantas infectadas. Para llevar a cabo una mejor caracterización del efecto de la metilación del DNA durante la infección por geminivirus, el triple mutante en *DRM1, DRM2* y *CMT3* de *A. thaliana* fue infectado con TYLCV o BCTV, y los niveles de DNA viral se midieron por separado en hojas apicales y en tejido floral. Los resultados no mostraron diferencias significativas al comparar plantas mutantes y silvestres infectadas con TYLCV. Sin embargo, sí observamos un aumento significativo en la cantidad de BCTV en el tejido floral de plantas mutantes, y un importante aumento en hojas apicales, aunque no significativo, debido a la alta variabilidad existente entre plantas diferentes. En conjunto, estos resultados sugieren que la inactivación de la maquinaria de metilación de mantenimiento en secuencias CG y CHG, dirigida por MET1 y CMT3 respectivamente, no interfiere en la acumulación de TYLCV en *A. thaliana*, del mismo modo ocurre en plantas deficientes en eliminación de la metilación por medio de ROS1. Además, los niveles de DNA de TYLCV no son significativamente diferentes en hojas o tejido floral de plantas silvestres y mutantes en *DRM1, DRM2* y *CMT3*. Sin embargo, los niveles de DNA de BCTV están incrementados en dichas plantas triple mutante, tanto en tejido floral como en hojas apicales.

Por otro lado, se llevaron a cabo ensayos en *N. benthamiana* con el objetivo de estudiar la importancia biológica de la metilación del DNA en la infección por TYLCV. Para ello, se empleó una tecnología consistente en el uso de un vector vírico (TRV, *Tobacco rattle virus*) para la inducción del silenciamiento génico a nivel post-transcripcional de un gen de interés, denominada VIGS (*Virus-induced gene silencing*). Mediante VIGS se indujo el silenciamiento de *NbMET1, NbCMT3* o *NbROS1* en plantas de *N. benthamiana*, las cuales fueron infectadas por agroinoculación con TYLCV. Se procesaron las hojas apicales de estas plantas para la extracción de RNA y DNA con el objetivo de determinar por RT-qPCR el grado de silenciamiento alcanzado en estos genes, y medir por qPCR los niveles de DNA viral. La acumulación de transcritos fue reducida a niveles por debajo del 30% para *NbMET1* y *NbCMT3*, y por debajo del 55% para *NbROS1*. Adicionalmente demostramos que el silenciamiento de *NbMET1* y *NbCMT3* induce hipometilación de secuencias repetidas del genoma de *N. benthamiana (NbHRS60)*. La acumulación de TYLCV fue similar en plantas silenciadas y plantas control no silenciadas, lo que indicaba que la desregulación del proceso de metilación en CG y CHG mediado por NbMET1 y NbCMT3, o de desmetilación mediado por NbROS1, no alteraba los niveles de ambos geminivirus en *N. benthamiana*.

Durante el análisis de los niveles de DNA viral en *N. benthamiana*, descubrimos que el silenciamiento de *NbMET1* o *NbCMT3* inducía mayor niveles de acumulación del gen ribosomal *Nb25S*. Este gen es usado por varios grupos como normalizador para la cuantificación por qPCR de la cantidad de geminivirus, por lo que concluimos que no era un gen válido para dicha normalización bajo el silenciamiento de la maquinaria de metilación. Tal observación impulsó la realización de nuevos experimentos para una mejor caracterización de este fenómeno, cuyos resultados son presentados en el capítulo 3.

Capítulo 3. Caracterización del DNA ribosomal (rDNA) en plantas deficientes en la metilación del DNA.

Previamente habíamos observado que la detección mediante qPCR del gen ribosomal *Nb25S* es alterada en plantas de *N. benthamiana* con reprimidos niveles de expresión de *NbMET1* y *NbCMT3*, sugiriendo que se produce un aumento en la cantidad de DNA de este gen ribosomal en la planta. En plantas existen dos unidades de DNA ribosomal (rDNA), ambas formadas por múltiples copias en tándem, pero localizadas en posiciones diferentes dentro de los cromosomas. La unidad subtelomérica *Nb45S* está formada por tres genes ribosomales: *Nb18S, Nb5.8S* y *Nb25S*; la unidad pericentromérica *Nb5S* sólo está formada por dicho gen. Trabajos de otros grupos han demostrado en levaduras, mosca, mamíferos y plantas que el rDNA es inestable y puede aumentar o disminuir su número de copias en determinadas situaciones. En levaduras, mosca y mamíferos se ha observado la existencia de rDNA en su forma episomal (extracromosómica y circular), cuya acumulación aumenta en mutantes de levadura de desacetilación de histonas, y en mutantes de metilación del DNA en mamíferos. Se ha sugerido que la formación de moléculas episomales tiene lugar mediante recombinación intramolecular entre copias de rDNA. En base a estos trabajos y nuestros resultados, hipotetizamos que la represión de la maquinaria de metilación del DNA de *N. benthamiana* induce inestabilidad genómica y un aumento en la cantidad de rDNA, y que dicho incremento podría darse mediante la generación de moléculas episomales. Para investigar esta hipótesis, silenciamos mediante VIGS los niveles de expresión de *NbMET1* y *NbCMT3*, y medimos por qPCR los niveles de DNA de los cuatro genes ribosomales que forman las unidades *Nb45S* y *Nb5S* (*Nb25S, Nb18S, Nb5.8S* y *Nb5S*. Los resultados mostraron que los niveles de rDNA correspondientes a los tres genes de la unidad *Nb45S* fueron más elevados en plantas silenciadas en *NbMET1* (alrededor de 4 veces más) y en *NbCMT3* (alrededor de 2 veces más), cuando comparados con plantas control inoculadas con el vector vacío. En cambio, no hubo tal aumento para el gen *Nb5S*. Dicho análisis fue realizado por qPCR usando dos genes normalizadores diferentes, *NbACT* y *NbEF1α*, obteniéndose datos similares con ambos. Adicionalmente, se midieron los niveles de rDNA (*25S* y *18S*) en plantas de *A. thaliana* deficientes en *MET1* (mutante *met1-3*) y observamos niveles más altos de *25S*, pero no de *18S*, en plantas *met1-3*, comparado con plantas Col-0.

Estos resultados sugieren que la desregulación de la metilación del DNA dirigida por NbMET1 y NbCMT3 provoca inestabilidad genómica en la región del rDNA correspondiente a la unidad *Nb45S*, induciendo un incremento de la cantidad de DNA de *Nb25S, Nb18S* y *Nb5.8S*. Para

confirmar dicha observación, se determinó la cantidad de rDNA en plantas control y plantas silenciadas en *NbMET1* o *NbCMT3*, mediante Southern blot y el uso una sonda marcada con digoxigenina correspondiente a una región de *Nb25S*. El resultado no mostró un aumento de la señal de rDNA en plantas silenciadas.

Con el objetivo de determinar la presencia de moléculas de rDNA episomales en plantas deficientes en la maquinaria de metilación del DNA, realizamos un Southern blot para detectar moléculas episomales mediante el uso de la misma sonda previamente utilizada y generada con la secuencia parcial de *Nb25S*. Esta metodología no nos permitió detectar dichas moléculas en plantas de *N. benthamiana* control o silenciadas en *NbMET1* o *NbCMT3.* Como última aproximación, seguimos una estrategia distinta consistente en el uso de RCA (*rolling circle amplification*), una técnica que permite amplificar DNA circular con muy alta eficiencia. Nuevamente, no se observó la presencia de moléculas circulares episomales. Para mejorar la capacidad de detección, combinamos RCA seguido de la cuantificación de las moléculas circulares amplificadas por qPCR, pero no detectamos la presencia de moléculas episomales para *Nb25S* en plantas cuya expresión de *MET1* había sido silenciada. Previamente se demostró que la combinación de las técnicas RCA y qPCR es válida para nuestro objetivo, ya que se detectó por qPCR la amplificación mediada por RCA de moléculas de geminivirus (DNA circular). Por tanto, estos datos sugieren que el silenciamiento de las metiltransferasas, y por tanto la inducción de hipometilación del genoma de *N. benthamiana*, alteran la detección y cuantificación por qPCR del rDNA.

Ya que la mayoría de copias de rDNA son transcripcionalmente reguladas por metilación del DNA y otras modificaciones epigenéticas, hipotetizamos que el silenciamiento de la metiltransferasa NbMET1 induciría la expresión del rDNA. Para testar dicha hipótesis, analizamos mediante RT-PCR, RT-qPCR y cuantificación directa del rRNA, la expresión de los genes ribosomales de *N. benthamiana* que forman parte de la unidad *Nb45S*, tanto del RNA precursor como de los transcritos maduros correspondientes a los genes *Nb25S, Nb18S* y *Nb5.8S*. No observamos mayor expresión del rDNA (RNA precursor o transcritos procesados) en plantas silenciadas en *NbMET1* al ser comparadas con plantas control no silenciadas, demostrando que la desregulación de la metilación CG mediada por esta enzima no induce la activación transcripcional de la unidad *Nb45S*.

Capítulo 4. No se encontró evidencia de transmisibilidad por semilla de BCTV y CaLCuV en *Arabidopsis*.

Tradicionalmente se ha considerado que los geminivirus sólo son transmitidos horizontalmente entre plantas vecinas a través de insectos vectores como la mosca blanca. Sin embargo, estudios recientes han mostrado que algunos geminivirus como TYLCV o BCTV también son transmitidos a la descendencia de plantas infectadas (plantas hospedadoras naturales) a través de las semillas. Este hallazgo tiene una especial importancia para un adecuado diseño de estrategias de control de enfermedades geminivirales en el campo y tiene un claro impacto en

investigación básica. El objetivo de este capítulo es evaluar la transmisión de los geminivirus por semilla a la descendencia de plantas de *A. thaliana* infectadas, focalizando el análisis en el begomovirus CaLCuV y el curtovirus BCTV.

Para evaluar la trasmisión del geminivirus a través de semilla, llevamos a cabo tres experimentos independientes en los cuales se infectaron plantas de *A. thaliana* con BCTV o CaLCuV. Para posteriores análisis, se eligieron aquellas plantas infectadas con BCTV o CaLCuV que produjeron semillas y mostraron mayor acumulación viral en tejido floral (5 plantas fueron seleccionadas para cada geminivirus). Por cada planta seleccionada, se escogieron 250 descendientes (25 días de edad) y se agruparon en 5 grupos (pools) de 50 plántulas cada uno y se determinaron los niveles virales mediante qPCR. Como control negativo, un total de 450 plántulas pertenecientes a la descendiencia de plantas no infectadas, fueron repartidas en 9 grupos y analizadas por qPCR. Para CaLCuV, se detectaron niveles de DNA viral similares en los grupos correspondientes a la descendencia de plantas infectadas y no infectadas. En cambio, para BCTV se detectaron niveles virales más altos en algunos grupos provenientes de plantas infectadas. Sin embargo, tras la amplificación por RCA y cuantificación por qPCR del DNA de dichos grupos de plantas, no se observó un incremento de los niveles de DNA de BCTV. Adicionalmente, para confirmar que BCTV no es transmitido por semilla en *A. thaliana*, se evaluó el desarrollo de síntomas y por tanto la presencia de geminivirus en plantas adultas (7 semanas de edad) correspondientes a la descendencia de plantas infectadas con BCTV. Tras analizar un total de 240 plantas, no se apreciaron síntomas de enfermedad viral en ninguna de ellas. En conjunto, estos datos indican que BCTV y CaLCuV no son transmitidos por semilla a la descendencia de plantas de *A. thaliana* infectadas, debido a que no se detectó acumulación en la siguiente generación.

Appendix I and II

Appendix I

Piedra-Aguilera et al., 2019. Scientific Reports. Vol. 9. Article number: 2863

Appendix II

Rodríguez-Negrete et al., 2013. New Phytologist 199, 464-475