



**INSTITUTO POTOSINO DE INVESTIGACIÓN
CIENTÍFICA Y TECNOLÓGICA, A.C.**

POSGRADO EN CIENCIAS EN BIOLOGIA MOLECULAR

**Determinación experimental del código de unión
específica al ADN de proteínas iniciadoras de la
replicación de geminivirus.**

Tesis que presenta
Angélica Romero Osorio

Para obtener el grado de
Doctora en Ciencias en Biología Molecular

Codirectores de la Tesis:
Dr. Gerardo Rafael Argüello Astorga
Dr. Guillermo Pastor Palacios

San Luis Potosí, S.L.P., Agosto del 2022



IPICYT

Constancia de aprobación de la tesis

La tesis “**Determinación experimental del código de unión específica al ADN de proteínas iniciadores de la replicación de geminivirus**” presentada para obtener el Grado de Doctora en Ciencias en Biología Molecular fue elaborada por **Angelica Romero Osorio** y aprobada el **16 de agosto de 2022** por los suscritos, designados por el Colegio de Profesores de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Dr. Gerardo Rafael Argüello Astorga
Codirector de la tesis

Dr. Guillermo Pastor Palacios.
Codirector de la tesis

Dr. Samuel Lara González
Miembro del Comité Tutorial

Dr. Roberto Ruiz Medrano
Miembro del Comité Tutorial



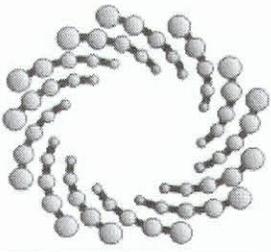
Créditos Institucionales

Esta tesis fue elaborada en el Laboratorio de Biología Molecular de Plantas de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la codirección de los doctores Gerardo Rafael Argüello Astorga y Guillermo Pastor Palacio.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología (**No.242730**) y del Instituto Potosino de Investigación Científica y Tecnológica, A. C.

Los proyectos de investigación descritos en esta tesis fueron financiados con recursos otorgados al Dr. Gerardo Rafael Argüello Astorga por el CONACYT (CONACYT-Fondos Sectoriales-SEP **CB-2015-01-257136**)

Agradecimiento especial al Laboratorio Nacional de biotecnología agrícola, médica y ambiental (LANBAMA), por los servicios otorgados.



IPICYT

Instituto Potosino de Investigación Científica y Tecnológica, A.C.

Acta de Examen de Grado

La Secretaría Académica del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 132 del Libro Primero de Actas de Exámenes de Grado del Programa de Doctorado en Ciencias en Biología Molecular está asentado lo siguiente:

En la ciudad de San Luis Potosí a los 30 días del mes de agosto del año 2022, se reunió a las 16:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

Dr. Samuel Lara González	Presidente	IPICYT
Dr. Guillermo Pastor Palacios	Secretario	UPIIG, IPN
Dr. Gerardo Rafael Argüello Astorga	Sinodal	IPICYT
Dr. Roberto Ruiz Medrano	Sinodal externo	CINVESTAV

a fin de efectuar el examen, que para obtener el Grado de:

DOCTORA EN CIENCIAS EN BIOLOGÍA MOLECULAR

sustentó la C.

Angélica Romero Osorio

sobre la Tesis intitulada:

Determinación experimental del código de unión específica al ADN de proteínas iniciadoras de la replicación de geminivirus

que se desarrolló bajo la codirección de

Dr. Gerardo Rafael Argüello Astorga
Dr. Guillermo Pastor Palacios (UPIIG, IPN)

El Jurado, después de deliberar, determinó

APROBARLA

Dándose por terminado el acto a las 18:15 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe la Secretaría Académica del Instituto.

A petición de la interesada y para los fines que a la misma convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 30 días del mes de agosto de 2022.


Dra. Lina Raquel Riego Ruiz
Secretaría Académica




Mtra. Ivonne Lizette Cuevas Vélez
Jefa del Departamento del Posgrado

IPICYT
SECRETARÍA ACADÉMICA
INSTITUTO POTOSINO DE
INVESTIGACIÓN CIENTÍFICA
Y TECNOLÓGICA. A.C.

Dedicatorias

Dedicado a todos los miembros de mi familia por siempre apoyarme y alentarme a seguir adelante.

Agradecimientos

A **CONACYT** por brindarme una beca para la realización de un posgrado.

Al **IPICYT** por su apoyo durante mi estancia

Al **Departamento de Posgrado y a los coordinadores del Posgrado** por su ayuda durante mi estancia.

Al **Dr. Argüello Astorga** por aceptarme y guiarme durante al doctorado.

Al **Dr. Pastor Palacios** por su apoyo, observaciones y amistad durante la realización del proyecto.

Al comité tutorial: **Dr. Samuel Lara González y Dr. Roberto Ruiz Medrano** por sus observaciones y aportaciones durante el desarrollo del trabajo.

Al **Dr. Bernardo Bañuelos Hernández**, por sus consejos como pionero en el trabajo con ToMoV, y por sus comentarios para mejorar una versión previa de la tesis.

Al biólogo **Salvador Ambriz Granados** por enseñarme y guiarme en el laboratorio, además de brindarme su amistad.

Al **LANBAMA**, en especial a Verónica Zarate por su especial apoyo en la secuenciación y la amistad brindada.

A **Rosy Castillo, Mireya Sanchez y la Sra Rosita**, por brindarme su amistad y hacerme sentir como en casa.

A mis compañeros de laboratorio 1, **Irene, Benita, Nadia, Tere, Nohemi, Norma, Karen, Marce, Aaron, Emanuel**, por su apoyo en el laboratorio haciendo más amenos aquellos días que eran mas malos que buenos.

Y, por último, pero no menos importante a **mamá, papá, Erika, Mayra, Pamela, Erick y Rene** por siempre estar conmigo y por su apoyo incondicional.

Contenido

Constancia de aprobación de la tesis	ii
Créditos institucionales	iii
Acta de examen	iv
Dedicatorias	v
Agradecimientos	vi
Lista de tablas	viii
Lista de figuras	ix
Abreviaturas	xi
Resumen	xiii
Abstract	xiv
Capítulo 1. Determinación experimental al código de unión específica al ADN de proteínas iniciadores de la replicación de geminivirus	
Introducción	1
Material y métodos	20
Resultados	26
Discusión y conclusión	59
Bibliografía	81
Capítulo 2- Explorando el origen del ancestro no-begomoviral del linaje del <i>Squash leaf curl virus</i>	67
Conclusiones	80
Referencias	81
Anexos	
Anexo 1: Artículo Publicado	89
Anexo 2: Material Suplementario	90
Anexo 3:	
Figura S1- Ilustración de un experimento independiente con ToMoV y los mutantes derivados.	94
Figura S2- Productos de PCR de los extractos de DNA de 35 plantas del experimento ilustrado en la figura S1.	95

Lista de tablas

Tabla 1. Miembros de la familia Geminiviridae	3
Tabla 2. Resumen de las proteínas de los begomovirus.	7
Tabla 3. DNA-A mutantes en los codones que corresponde a los DEs.	33
Tabla 4. Componentes A dobles mutantes (DEs/ iterones cognados)	39
Tabla 5. Clonas infectivas de componentes A mutantes y silvestre.	41
Tabla 6. Resumen de las clonas infectivas del DNA-B construidas en este trabajo.	47
Tabla 7. Resumen de las combinaciones de <i>A. tumefaciens</i> GV3101 utilizadas en los ensayos de infección con ToMoV y sus derivados mutantes.	48
Tabla 8. Reporte de sintomatología de plantas infectadas	49
Tabla S1. Oligonucleótidos mutagénicos para generar el sitio <i>NgoMIV</i>	75
Tabla S2. Casetes sintéticos con DEs diferentes a ToMoV	75
Tabla S3. Oligonucleótidos para mutaciones sitio-dirigidas de DEs 69 y 71 diferentes a ToMoV	76
Tabla S4. Mega-oligonucleótidos con iterones difentes a ToMoV	77

Lista de figuras

Capítulo 1

Figura 1. Geminivirus	1
Figura 2. Organización genómica e insectos transmisores de los virus pertenecientes a los diferentes géneros de geminivirus	4
Figura 3. Organización genómica de un begomovirus bipartita	6
Figura 4. Ciclo infectivo de un begomovirus monopartita	8
Figura 5. Mecanismo de replicación por círculo rodante	10
Figura 6. Representación esquemática de la proteína de inicio de la replicación.	11
Figura 7. Organización de los sitios de unión (iterones) y de corte (sitio “Nick”) de la proteína Rep en la región común de ambos componentes virales.	12
Figura 8. Auto-replicación en células de <i>N. tabacum</i> e infección de plantas de tabaco de los componentes genómicos híbridos de ToMoV.	14
Figura 9- Modelo de la estructura 3D predicha <i>in silico</i> del dominio de unión al ADN de la proteína Rep de SiMSinV.	15
Figura 10- Alineamiento de la región N-terminal de proteínas Rep agrupadas por la similitud de los aminoácidos 5, 8, 10, 69 y 71.	16
Figura 11- Modelo Predictivo	28
Figura 12. Introducción del sitio <i>NgoMIV</i> en la secuencia codificante del gen <i>Rep</i> en el componente A de ToMoV	29
Figura 13. Estrategia para el reemplazo de 37 pb del inicio del gen <i>Rep</i> mediante casetes sintéticos mutantes	31
Figura 14. Clonación de casetes sintéticos en componente A de ToMoV	32
Figura 15. Alineamiento de los primeros 80 aa de las proteínas Rep mutantes y silvestre de ToMoV	34
Figura 16. Ensayos de auto-replicación de componentes mutantes de ToMoV en protoplastos de tabaco NT1	35
Figura 17. Diseño de oligonucleótidos extensos.	36
Figura 18. Construcción de los DNA-A “doble mutantes”.	37
Figura 19. Componente M3A3	38
Figura 20. Mapa físico de las clonas infectivas de DNA-A de ToMoV-A mutante o silvestre.	41
Figura 21. Diagrama ilustrativo de ensayos de auto-replicación en discos de hoja	42
Figura 22. Ensayos de auto-replicación en discos de hoja	44
Figura 23. Mapa físico del DNA-B1 mutante tipo BCTV.	45
Figura 24. Mapa físico de la clona infectiva B1P1	46
Figura 25. Ensayo de inoculación de componentes A mutantes más componente B silvestre de ToMoV	48
Figura 26. Ensayos de infección de <i>N. benthamiana</i> .	50
Figura 27. Detección de componentes virales en ensayos de infección con componentes A mutantes y B quiméricos.	51
Figura 28. Alineamientos de secuencias de productos de PCR derivados de plantas infectadas que corresponden al componente genómico A	51
Figura 29. Alineamientos de secuencias de productos de PCR derivados de los componentes genómicos B aislados de plantas infectadas.	53
Figura 30. Análisis de secuencia con el software SnapGene®, de los productos de PCR recuperados de plantas infectadas con ToMoV.	53

Figura 31. Análisis de secuencia con el software SnapGene®, de los PCRs recuperados de plantas infectadas con el mutante ToMoV de tipo PHYVV y ACMV	54
Figura 32. Análisis con el software SnapGene®, de las secuencias virales recuperadas de plantas infectadas con el mutante ToMoV de tipo OkYMMV y TGMV	55
Figura 33. Análisis con el software SnapGene® de las secuencias virales recuperadas de plantas infectadas con el mutante ToMoV de tipo BCTV	56
Figura 34. Modelo estructural del extremo N-terminal de la proteína Rep de ToMoV	62
Figura 35. Modelo estructural <i>in silico</i> de la interacción proteína Rep-DNA de TGMV and CaLCuV	65

Capítulo 2

Figura 2.1. Relaciones filogenéticas de los miembros del linaje del SLCV con otros geminivirus, basadas en el dominio N-terminal (1-162).	69
Figura 2.2. Análisis BLASTp con los primeros 120 residuos aa de la proteína Rep de Tomato severe leaf curl virus (ToSLCV).	70
Figura 2.3. Alineamiento del dominio endonucleolítico de la proteína Rep de ToSLCV con una proteína Rep de un virus de invertebrados	72
Figura 2.4. Relaciones filogenéticas de la proteína AC4* del linaje del SLCV con proteínas C4/AC4 de otros geminivirus.	74
Figura 2.5. Alineamiento del dominio N-terminal de la proteína AC4* con sus análogos posicionales C4/AC4 de diversos geminivirus.	75
Figura 2.6. Representación simplificada de los ORFs en sentido complementario que codifican a Rep y otras proteínas en varios geminivirus.	79

Abreviaturas

ADNss	ADN circular de cadena sencilla
ACMV	<i>African cassava mosaic virus</i>
BCTIV	<i>Beet curly top Iran virus</i>
BCTV	<i>Beet curly top virus,</i>
BGMV	<i>Bean golden mosaic virus,</i>
Cb	Carbenicilina
CCDaV	<i>Citrus chlorotic dwarf associated virus.</i>
CdTV	<i>Chino del tomate virus</i>
CP	Proteína de la cápside
DEs/SPDs	Determinantes de especificidad (Rep “Specificity Determinants”)
dsDNA	DNA de doble cadena (double stranded DNA)
EcmlV	<i>Euphorbia caput-medusae latent virus</i>
ECSV	<i>Eragrostis curvula streak virus</i>
EuMV	<i>Euphorbia mosaic virus</i>
Gm	Gentamicina
GRBV	<i>Grapevine red blotch virus</i>
IRD	Dominio-relacionado al iterón (Iteron-Related Domain)
Km	kanamicina
MP	Proteína involucrada en el movimiento intercelular del virus.
MSV	<i>Maize streak virus</i>
NM	Nuevo Mundo
NSP	Nuclear Shuttle Protein (proteína lanzadera nuclear)
OriV	origen de replicación viral
PHYVV	<i>Pepper huasteco yellow vein virus</i>
RCR	Replicación por Círculo Rodante (Rolling-circle replication)
REn	Proteína potenciadora de la replicación
Rep	Proteína iniciadora de la replicación viral
Rf	Rifampicina
SiMSinV	<i>Sida mosaic Sinaloa virus</i>

ssDNA	DNA de cadena sencilla (single-strand DNA)
TCTV	<i>Turnip curly top virus</i>
TGMV	<i>Tomato golden mosaic virus</i>
ToMoTV	<i>Tomato mottle Taino virus</i>
ToMoV	<i>Tomato mottle virus</i>
TPCTV	<i>Tomato pseudo-curly top virus,</i>
TrAP	Proteína transactivadora, proteína C2/A(C2)
TYLCSV	<i>Tomato yellow leaf curl Sardinia virus</i>
TYLCV	<i>Tomato yellow leaf curl virus</i>

Resumen

Determinación experimental del código de unión específica al ADN de proteínas iniciadoras de la replicación de geminivirus

Los geminivirus constituyen la familia más extensa de virus de plantas; poseen un genoma circular de DNA de cadena sencilla (ssDNA) y codifican una proteína iniciadora de la Replicación por Círculo Rodante (RCR), denominada "Rep". Estas proteínas reconocen y unen con alta afinidad elementos cortos repetidos, denominados "iterones", que se ubican en el origen de replicación viral. En los geminivirus se han descrito decenas de iterones diferentes, y existe un evidente interés por comprender el código molecular que determina la preferencia de Rep por iterones específicos. Mediante una combinación de enfoques teóricos y experimentales, nuestro grupo ha logrado identificar cinco potenciales Determinantes de Especificidad (DEs), esto es, residuos aminoacídicos que podrían determinar la preferencia de Rep por secuencias de DNA específicas. En el caso del geminivirus bipartita ToMoV (virus del moteado del tomate) esos cinco residuos de Rep se ubican en las posiciones 5, 8, 10, 69 y 71. El objetivo del presente estudio fue examinar experimentalmente la hipótesis antes mencionada. A partir de un modelo predictivo de presuntos pares cognados RepDEs-Iterón se construyeron 8 componentes genómicos A (DNA-As) de ToMoV con mutaciones en los codones de los DEs hipotéticos. Estos DNA-As mutantes fueron examinados en ensayos de auto-replicación transitoria en protoplastos de tabaco NT1. Los DNA-A mutantes se replicaron a niveles muy inferiores al virus silvestre, y en algunos casos no se detectó DNA viral *de novo*, lo que sugiere que las proteínas con mutaciones en los residuos 5, 8, 10, 69 y 71, perdieron, parcial o totalmente, la capacidad de unirse a los iterones nativos de ToMoV. En una segunda fase, generamos DNA-As en los que los iterones de ToMoV fueron sustituidos por los presuntos sitios de unión de las proteínas Rep mutantes. Se evaluó la capacidad replicativa de cinco de esos DNA-As "doblemente mutados" de ToMoV, mediante ensayos de auto-replicación en discos de hoja, observándose la generación de DNA viral *de novo*, lo que indica que las proteínas Rep mutadas reconocieron a los iterones predichos por el modelo hipotético. Finalmente, se generaron componentes B mutantes, en los que los iterones de ToMoV fueron sustituidos por iterones idénticos a los correspondientes a los DNA-As mutantes, y se realizaron ensayos de agroinoculación de plantas de *N. benthamiana*, usando clones infectivos de cinco de los componentes genómicos A y B mutantes. Las plantas inoculadas presentaron síntomas visibles a partir del día 14 post-inoculación (dpi). Se extrajo DNA total de las hojas nuevas de 3 plantas de cada grupo experimental, y se amplificaron los segmentos con el origen de replicación y parte del gen *Rep* con iniciadores específicos. La secuenciación de los productos de PCR confirmó que los DNA virales en las plantas sintomáticas correspondían a los componentes A y B mutantes que fueron inoculados. Estas observaciones apoyan fuertemente la hipótesis general propuesta, y contribuyen de modo significativo al desciframiento del código que determina el reconocimiento específico de los iterones por las proteínas Rep de geminivirus.

PALABRAS CLAVE.

Begomovirus, Iterones, determinantes de especificidad replicativa, replicación por círculo rodante, interacciones DNA-proteína.

Abstract

Experimental determination of the code of specific binding to viral DNA of the rolling-circle replication initiation protein encoded by geminiviruses.

The family *Geminiviridae* is the largest and ecologically most successful family of plant viruses, its genomes consist of one or two small, circular ssDNA molecules encoding a few multifunctional proteins, including a rolling-circle replication (RCR) initiation protein, termed Rep, that shares several conserved amino acid sequence motifs with many other viral and plasmidic RCR proteins. The Rep protein recognizes and binds with high-affinity short repeat DNA sequences, denominated "iterons", which are sited close to the geminiviral origin of replication. Dozens of different iterons have been described; consequently, there exists a natural interest for knowing the molecular code that ruled the preference of Rep by specific iterons. Through a combination of theoretical and experimental approaches, our group has been able to (tentatively) identify a series of amino acid residues in the geminivirus Rep that probably determine its preference for specific iterons. We have identified five major potential Specificity Determinants (SDs) of the Rep DNA-binding domain, that is, aa residues that could determine the Rep preference for specific iterons. In the case of the bipartite geminivirus ToMoV (tomato mottle virus) those five residues have been mapped at positions 5, 8, 10, 69, and 71 of Rep. The aim of this study was to validate the aforesaid hypothesis. Based on a predictive model of postulated RepDEs-Iteron cognate pairs, 8 genomic components A (DNA-As) of ToMoV were constructed displaying mutations at the hypothetical DEs codons. These mutant DNA-As were examined in transient self-replication assays in tobacco protoplasts NT1. The mutant A-DNAs replicated at much more lower levels than wild-type viruses, hence suggesting that the mutated Rep partially or totally, lost the ability to bind the native ToMoV iterons. In a second phase, we generated DNA-As in which the ToMoV iterons were replaced by the putative binding sites of the mutant Rep proteins. The replicative capacity of five of these ToMoV-A "double mutants" was evaluated in self-replication assays on leaf discs. As a result, the generation of *de novo* viral DNA was observed, although the variability between independent agroinoculation assays was elevated. Finally, mutant B components were generated, in which the iterons of ToMoV-B were replaced by iterons identical to those corresponding to the mutant DNA-As, and were evaluated by agroinoculation assays of *N. benthamiana* plants, using infective clones of five of the mutant genomic components A and B. The inoculated plants presented visible symptoms at 14 days post-inoculation (dpi). Total DNA was extracted from the new leaves of plants of each experimental group and the segments with the origin of replication and part of the Rep gene were amplified with specific primers. The sequencing of the PCR products confirmed that viral DNAs in symptomatic plants corresponded to the mutant components A and B that were inoculated. These observations strongly support the proposed general hypothesis and contribute significantly to deciphering the code determining the specific recognition of iterons by geminivirus Rep proteins.

Keywords:

Begomovirus, Iterons, specificity determinants of replication, Rolling circle replication, DNA- protein interaction

I. INTRODUCCIÓN

Características generales de los geminivirus

Los geminivirus constituyen una familia viral muy extensa, con poco más de 500 especies reconocidas, de acuerdo con el International Committee on Taxonomy of Viruses (ICTV) [<http://www.ictvonline.org/virustaxonomy.asp>] (Fiallo-Olivé *et al.*, 2021; Zerbini *et al.*, 20017; Silva *et al.*, 2017). Estos virus infectan a una enorme variedad de plantas, tanto silvestres como cultivadas, que incluyen leguminosas, hortalizas, cereales, plantas ornamentales y productoras de fibras, y causan enormes pérdidas económicas en todas las regiones cálidas y templadas del mundo (Rojas *et al.*, 2005; Rojas *et al.*, 2018). Los geminivirus han sido recientemente incluidos en el phylum Cressdnaviricota (Krupovic *et al.*, 2020), un taxón que unifica a 7 familias de virus que poseen un genoma circular de DNA de cadena sencilla (ssDNA) y que codifican proteínas iniciadoras de la Replicación por Círculo Rodante (RCR), denominadas “Rep”. Los geminivirus se diferencian del resto de los virus conocidos por la peculiar estructura de su cápside, formada por dos icosahedros incompletos fusionados, que les da apariencia de partículas gemelas o “geminadas” al microscopio electrónico (Zhang *et al.*, 2001 Hesketh *et al.*, 2018, Hipp *et al.*, 2018) (Figura 1).

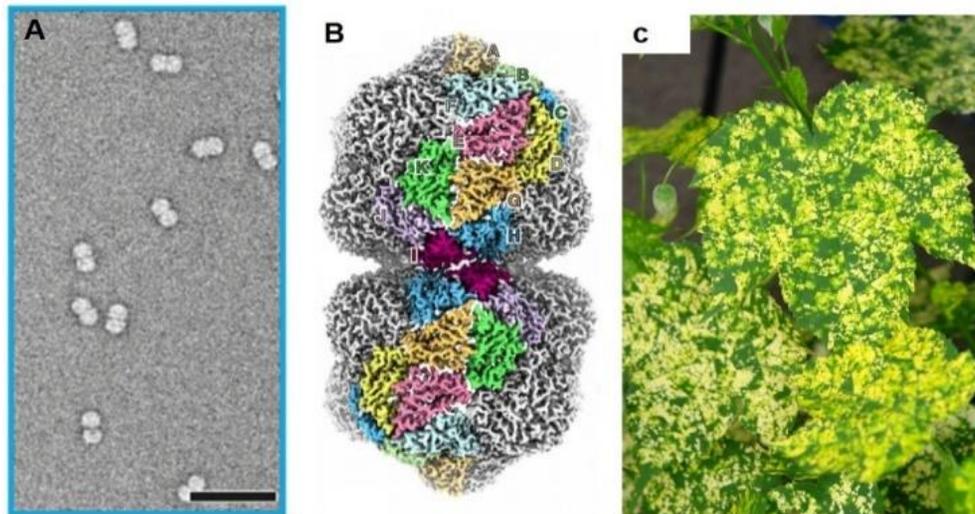


Figura 1. Geminivirus. (A). Microscopía electrónica de partículas geminadas. **(B).** Esquematización de la partícula geminada **(C).** Planta de abutilón infectada con *Abutilon mosaic virus*. Modificada de Hesketh *et al.*, 2018.

Su genoma se compone de una o dos moléculas circulares de ADN de cadena sencilla (ADNcs) de 2600 a 3000 nucleótidos de longitud (Jeske 2009). El reducido genoma de los geminivirus codifica pocas proteínas (de 4 a 8), por lo que dependen en gran medida de la maquinaria molecular del huésped para replicar su genoma y expresar sus genes (Hanley-Bowdoin *et al.*, 2013, Jeske *et al.*, 2009). Los geminivirus interfieren

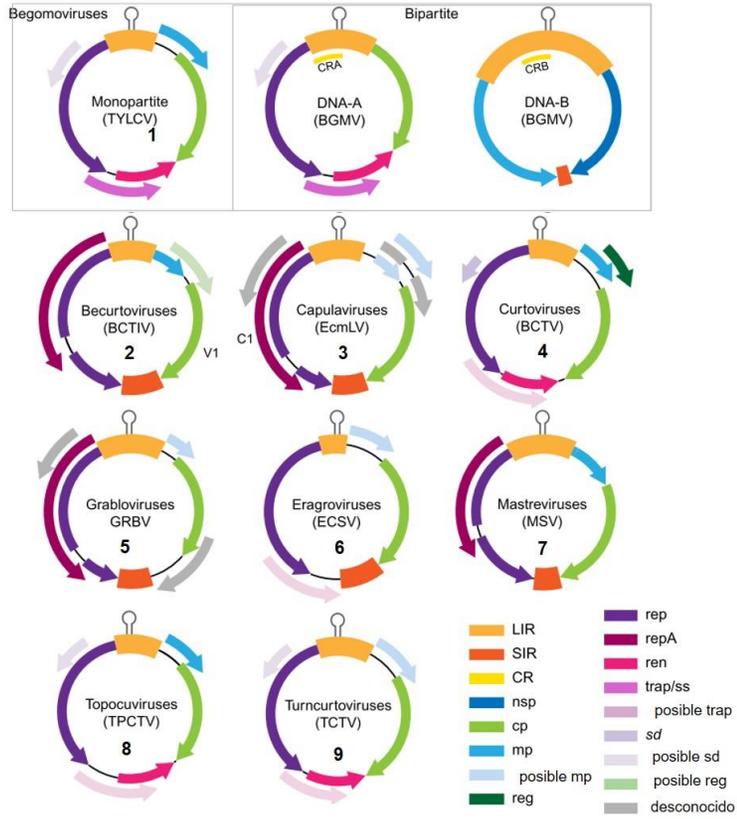
el ciclo celular de las células vegetales para favorecer su replicación, además de alterar el patrón de expresión génica y el tráfico macromolecular en el huésped, interfieren las vías de señalización celular y redirigen o bloquean las respuestas de defensa de la planta (Bisaro *et al.*, 2006, Buchmann *et al.*, 2009, Raja *et al.*, 2010). Adicionalmente, los geminivirus alteran la metilación del ADN y codifican varios supresores del silenciamiento que interfieren la producción de RNAs interferentes (Bisaro *et al.*, 2006, Paprotka *et al.*, 2011, Pumplin y Voinnet 2013; Vanitharani *et al.*, 2005.). Los miembros de la familia *Geminiviridae* se clasificaron, hasta muy recientemente, en nueve géneros, de acuerdo con su organización genómica, las plantas hospedadoras, y los insectos que los transmiten: *Becurtovirus*, *Begomovirus*, *Capulavirus*, *Curtovirus*, *Eragrovirus*, *Grablovirus*, *Mastrevirus*, *Topocuvirus*, y *Turncurtovirus* (Tabla 1 y Figura 2) (Navas-Castillo, J. 2011; Varsani *et al.*, 2014, 2017; Zerbini *et al.*, 2017). A fines del 2021 se describieron cinco nuevos géneros, con 1 a 4 especies cada uno: *Citlodavirus*, *Maldovirus*, *Mulcrilevirus*, *Opunvirus* y *Topilevirus* (Fiallo-Olive *et al.*, 2021). El género *Begomovirus* agrupa el mayor porcentaje (87%) de la familia, y sus miembros se caracterizan por ser transmitidos por la mosquita blanca (*Bemisia tabaci* Genn.) una plaga agrícola mundial (Brown *et al.*, 2015; De Barro *et al.*, 2011), poseer un genoma compuesto por una sola molécula de DNA (monopartita) o por dos moléculas diferentes (bipartita) y por infectar plantas dicotiledóneas (Brown *et al.*, 2015). Este género es el único, entre todos los geminivirus, que incluye especies con un genoma segmentado (bipartita).

Tabla 1. Miembros de la familia *Geminiviridae* (Modificada de Zhao *et al.*, 2019).

Familia	Genero	Número de especies	Especie representativa	Número de acceso
<i>Geminiviridae</i>	<i>Becurtovirus</i>	3	Beet curly top Iran virus	EU273818
	<i>Begomovirus</i>	409	Bean golden yellow mosaic virus	DNA-A: M88686 DNA-B: M88687
	<i>Capulavirus</i>	4	Euphorbia caput-medusae latent virus	KT214376
	<i>Curtovirus</i>	3	Beet curly top virus	M24597
	<i>Eragrovirus</i>	1	Eragrostis curvula streak virus	FJ665631
	<i>Grablovirus</i>	3	Grapevine red blotch virus	KC896623
	<i>Mastrevirus</i>	40	Maize streak virus	AF329878
	<i>Topocuvirus</i>	1	Tomato pseudo-curly top virus	X84735
	<i>Turncurtovirus</i>	2	Turnip curly top virus	GU456685

Los genomas de los geminivirus están sujetos a múltiples cambios, debido a sus altas tasas de mutación, recombinación, pseudo-recombinación, sinergismo con otros virus y la transmisión mediada por vectores (insectos), lo que impulsa el aumento de la diversidad viral (Rojas *et al.*, 2005; Czosnek *et al.*, 2017; Kumar y Chakraborty 2017).

A



B

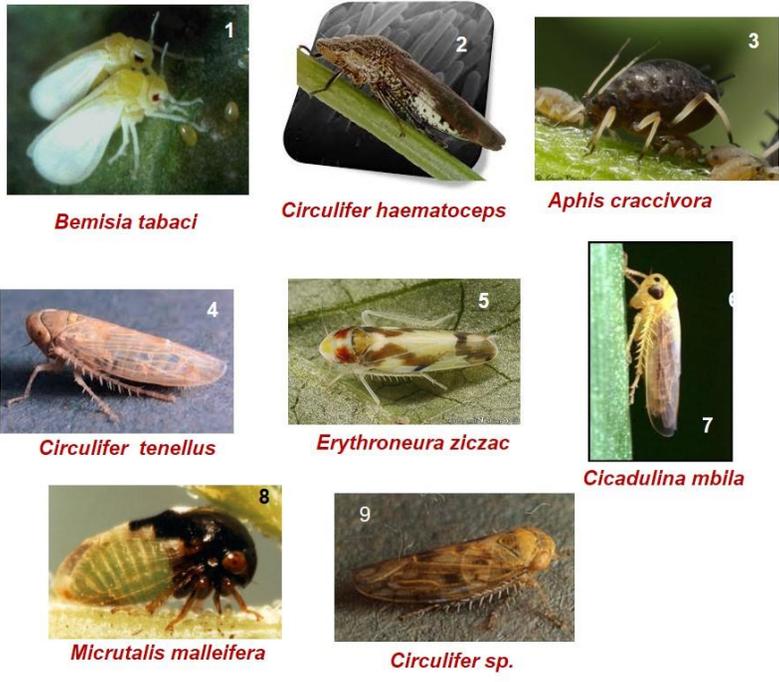


Figura 2. Organización genómica e insectos transmisores de los virus pertenecientes a 9 diferentes géneros de geminivirus. Abreviaturas: TYLCV, tomato yellow leaf curl virus, BGMV, bean golden mosaic virus, BCTIV, beet curly top Iran virus, EcmLV, Euphorbia caput-medusae latent virus, BCTV, beet curly top virus, GRBV, grapevine red blotch virus, ECSV, Eragrostis curvula streak virus, MSV, maize streak virus, TPCTV, tomato pseudo-curly top virus, TCTV, turnip curly top virus, CCDaV, citrus chlorotic dwarf associated virus. (Adaptado de Bernardo *et al.*, 2012)

Genoma de los begomovirus

Todos los begomovirus del continente americano (“Nuevo Mundo”, NM), con unas pocas excepciones (Melgarejo *et al.*, 2013; Romay *et al.*, 2016), y una parte menor pero significativa, de los del Viejo Mundo (VM), poseen un genoma compuesto por dos moléculas circulares denominadas DNA-A y DNA-B (Figura 3). El componente genómico A contiene cinco ORFs: 1) *AC1* o *Rep*, que codifica la proteína que inicia la replicación viral (*Rep*) (Ingham *et al.*, 1995, Singh *et al.*, 2008); 2) *AC2* o *Trap*, que codifica una proteína multifuncional (*TrAP*) que transactiva los genes tardíos y funciona como supresor de silenciamiento (Fondong, 2013; Vanitharani *et al.*, 2005); 3) *AC3* o *REn*, que produce una proteína (*REn*) que potencia la replicación viral, y actúa como supresor del silenciamiento transcripcional (Settlage *et al.*, 2005, Rodríguez-Gandarilla *et al.*, 2020); 4) *AC4*, que codifica un supresor del silenciamiento postranscripcional (Medina-Puche 2021); y 5) *AV1* o *CP*, que codifica la proteína de la cubierta (“coat”) o cápside (*CP*) (Sharma *et al.*, 2009; Teng *et al.*, 2010.). El componente genómico B, por otra parte, contiene dos genes que se transcriben en direcciones opuestas: 1) *BV1* o *NSP*, que da lugar a una proteína (*NSP*, Nuclear Shuttle Protein) que transporta el ADN viral del núcleo al citoplasma, y de éste al núcleo (Zhou *et al.*, 2011, Sanderfoot *et al.*, 1996); y 2) el gen *BC1* o *MP* que codifica la proteína *MP*, involucrada en el movimiento intercelular y sistémico del ADN viral (Fondong, 2013) (Tabla 2). La región no codificante que separa a los dos grupos de genes que se transcriben de modo divergente, contiene el origen de replicación (*Ori*) viral, localizado en un segmento de ~170 a 190 nt que es casi idéntico en ambos componentes genómicos (Hanley-Bowdoin *et al.*, 2013; Revington *et al.*, 1989). El segmento compartido se denomina “Región Común” y es el que hace posible la replicación coordinada de los dos componentes genómicos (Brown *et al.*, 2012; Jeske, 2009, Revington *et al.*, 1989.).

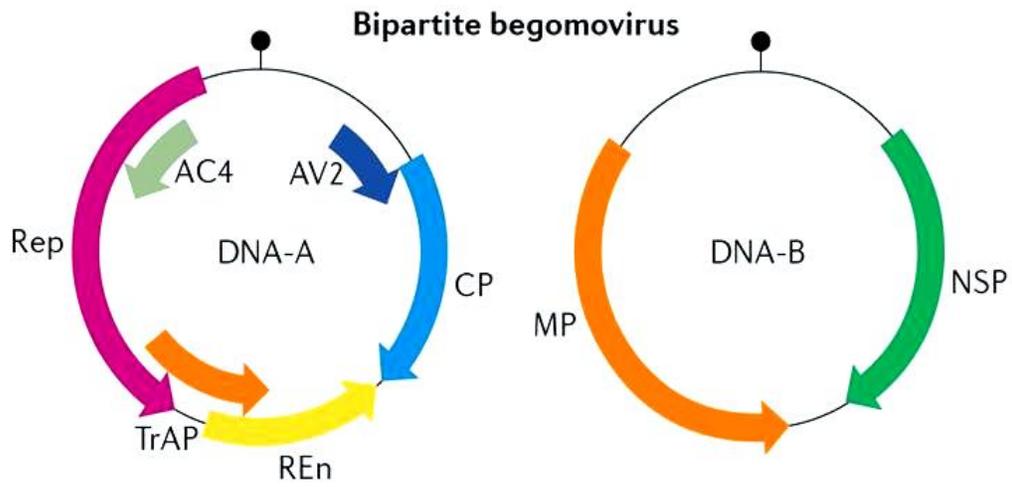


Figura 3. Organización genómica de un begomovirus bipartita (reproducido de Hanley Bowdoin *et al.*, 2013).

Tabla 2. Resumen de las proteínas codificadas de los begomovirus.

Gen	Proteína	Tamaño (aa)	Función
<i>AC1, Rep</i>	Rep	345-361	Inicia la replicación viral. Interfiere el ciclo celular
<i>AC2, TrAP</i>	TrAP	130	Activa la expresión de los genes tardíos. Suprime la respuesta de silenciamiento génico
<i>AC3, REn</i>	REn	132	Potencia la replicación
<i>AC4</i>	AC4	87	Supresor del silenciamiento génico.
<i>AV1, CP</i>	CP	254	Forma la cápside viral, determina la transmisión por el insecto vector.
<i>AV2</i>	AV2	72	Supresor del silenciamiento en virus del Viejo Mundo (VM)
<i>BV1, NSP</i>	NSP	300	Trafico de ssDNA viral entre el núcleo y el citoplasma (o viceversa)
<i>BC1, MP</i>	MP	300	Movimiento célula-célula y sistémico.

Ciclo infeccioso de los begomovirus

El ciclo de infección de los begomovirus comienza cuando la cápside del virión, introducido por la mosquita blanca virulífera, se desensambla, liberando el DNA viral monocatenario (ssDNA) en o cerca del núcleo. Las polimerasas del huésped sintetizan la cadena complementaria para generar un DNA de doble cadena (dsDNA) que es recubierto enseguida por nucleosomas, formando un minicromosoma (Donson *et al.*, 1984, Pilartz y Jeske 1992.) El intermediario de dsDNA es transcripcionalmente activo, y constituye el sustrato indispensable para la replicación por círculo rodante (RCR) del genoma viral (Laufs *et al.*, 1995, Bisaro 2009). Los mRNAs son producidos por la RNA polimerasa II del huésped, y su subsecuente traducción da origen a las proteínas virales “tempranas” (Singh *et al.*, 2007, Settlege *et al.*, 2005), entre las que destaca la proteína iniciadora de la replicación (Rep), esencial para el proceso de RCR (Gutierrez *et al.*, 2004, Shimada-Beltran *et al.*, 2007). El ssDNA monomérico producido en cada ciclo de RCR es convertido en DNA de doble cadena para volver a entrar en el ciclo de replicación (Hanley-Bowdoin *et al.*, 2013). La proteína Rep reprime la transcripción de

su propio gen, lo que al mismo tiempo incrementa la expresión del gen contiguo, AC2, que codifica la proteína activadora de la transcripción (TrAP). Esta proteína activa la expresión de los genes tardíos, que codifican a las proteínas de la cápside (CP) y de movimiento, NSP y MP (Hung y Petty 2001). En la última fase del ciclo de multiplicación viral cada uno de los componentes genómicos es empaquetado, por separado, en una cápside formada por 110 moléculas de la proteína CP, para formar nuevas partículas infecciosas (viriones) (Zhang *et al.*, 2003). La NSP también se une al DNA viral y lo mueve del núcleo al citoplasma (y viceversa) atravesando la membrana nuclear y, en asociación con la proteína MP, lo transporta además a las células vecinas a través de los plasmodesmos (Figura 4) (Hanley-Bowdoin *et al.*, 2013; Mengshi *et al.*, 2022).

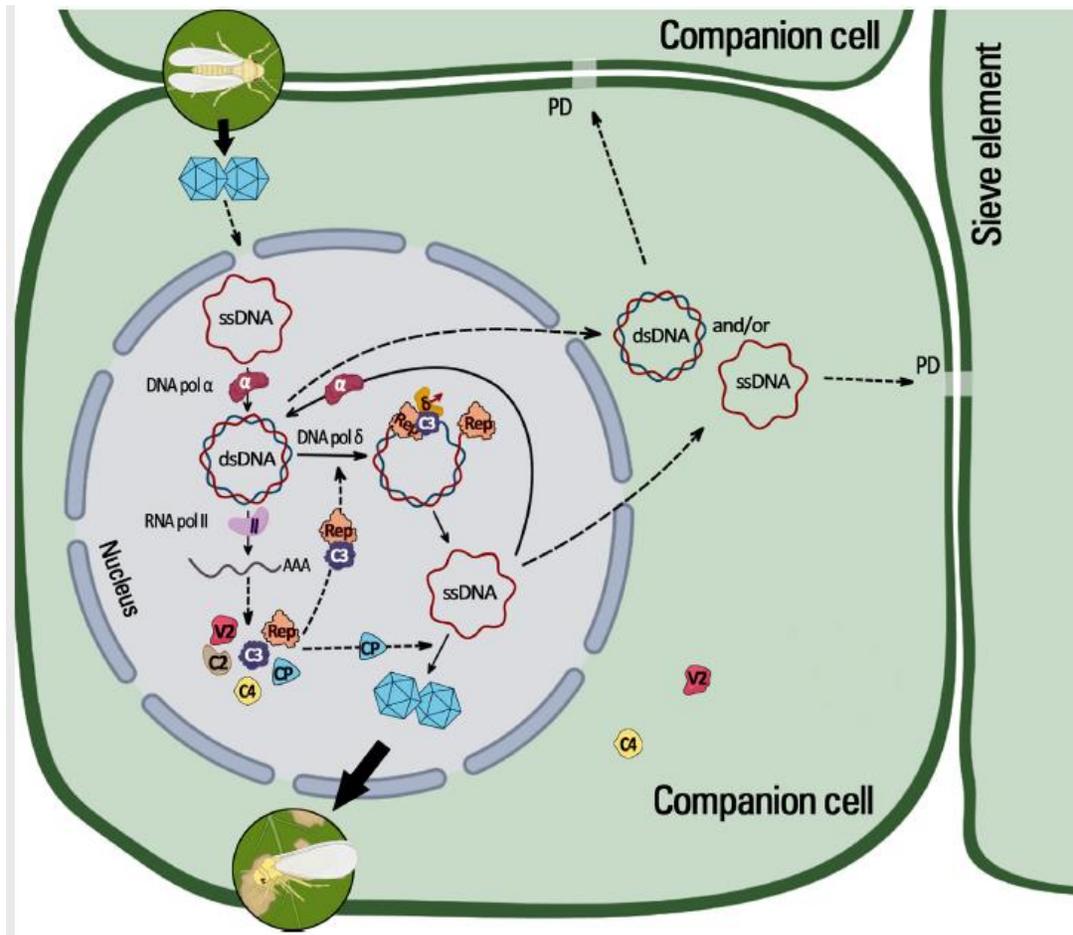


Figura 4. Ciclo infeccioso de un begomovirus monopartita, como TYLCV. La infección comienza cuando el insecto vector se alimenta de la planta y libera viriones dentro de una célula vegetal. La cápside se desensambla y el genoma viral liberado llega al núcleo, donde tiene lugar el proceso de replicación. En ese organelo, el ssDNA es convertido a un intermediario de dsDNA con la participación de la DNA polimerasa α . Los genes del dsDNA son transcritos por la RNA polimerasa II, lo que resulta en la producción de proteínas virales. La proteína Rep se une al origen de replicación e introduce un corte endonucleolítico ("nick") en la cadena del virión (+) para iniciar la replicación del vDNA por un mecanismo de Círculo Rodante,

para el cual la DNA polimerasa γ es requerida. El DNA viral recién sintetizado puede moverse intra- y extracelularmente, estableciendo una infección sistémica. Adicionalmente, el ssDNA viral puede ser encapsidado por CP y adquirido por los insectos vectores, y de ese modo ser transmitido a otras plantas. (reproducido de Mengshi *et al.*, 2022).

Replicación por Círculo Rodante (RCR)

El mecanismo de replicación por Círculo Rodante (RCR) es empleado por virus de DNA de cadena sencilla como los geminivirus (Lazarowitz *et al.*, 1992). El genoma viral de cadena sencilla se convierte a DNA de doble cadena (dsDNA) por la acción de la DNA pol α , y este dsDNA es la forma transcripcional y replicativamente activa (Jeske *et al.*, 2001; Preiss y Jeske, 2002), mediante la acción de factores y las DNA polimerasas del huésped (Saunders *et al.*, 1991; Shimada-Beltran y Rivera-Bustamante, 2007). Durante el inicio del proceso, la proteína asociada a la replicación viral (Rep) se une a la forma replicativa cerca del origen de la replicación (oriV, donde induce la formación de una estructura de tipo “tallo-asa” a través de interacciones todavía no bien comprendidas (Argüello-Astorga *et al.*, 1994a, 1994b; Rizvi *et al.*, 2015), pero que podrían involucrar la interacción con factores transcripcionales del huésped, que favorecerían la formación de una horquilla (en realidad, una estructura “cruciforme”), lo que expondría el “asa” de la horquilla de la cadena + en forma de ssDNA, la que constituye el sustrato endonucleolítico de la proteína Rep. Esta última introduce un corte (“nick”) en la cadena del virión (“plus”) del vDNA, entre el nucleótido 7 y 8 de la secuencia conservada 5'-TAATATT*AC-3', ubicada en el ápice del elemento tipo “tallo-asa” (Laufs *et al.*, 1995). El corte endonucleolítico genera un extremo 3' libre, que actúa como “iniciador” para la síntesis de una copia de la molécula circular de DNA viral por acción de las DNA polimerasas de la célula. En este proceso Rep permanece unida covalentemente al extremo 5' del DNA cortado. Durante la replicación, la cadena original es desplazada por la cadena positiva recién sintetizada. Cuando se completa la replicación, Rep vuelve a cortar en el asa regenerada, y une catalíticamente los extremos generados (ligación), permitiendo que la cadena sencilla del DNA inicial se libere de la forma replicativa (Figura 5) y se forme un nuevo intermediario de dsDNA.

En el caso de los geminivirus, un paso fundamental antes de iniciar la RCR, es el reconocimiento y unión específica a cortas secuencias de dsDNA, denominados “iterones” por repetirse varias veces en la región del origen de replicación viral (Argüello-Astorga *et al.*, 1994b; Castellanos *et al.*, 1999; Jeske *et al.*, 2001; Singh *et al.*, 2008).

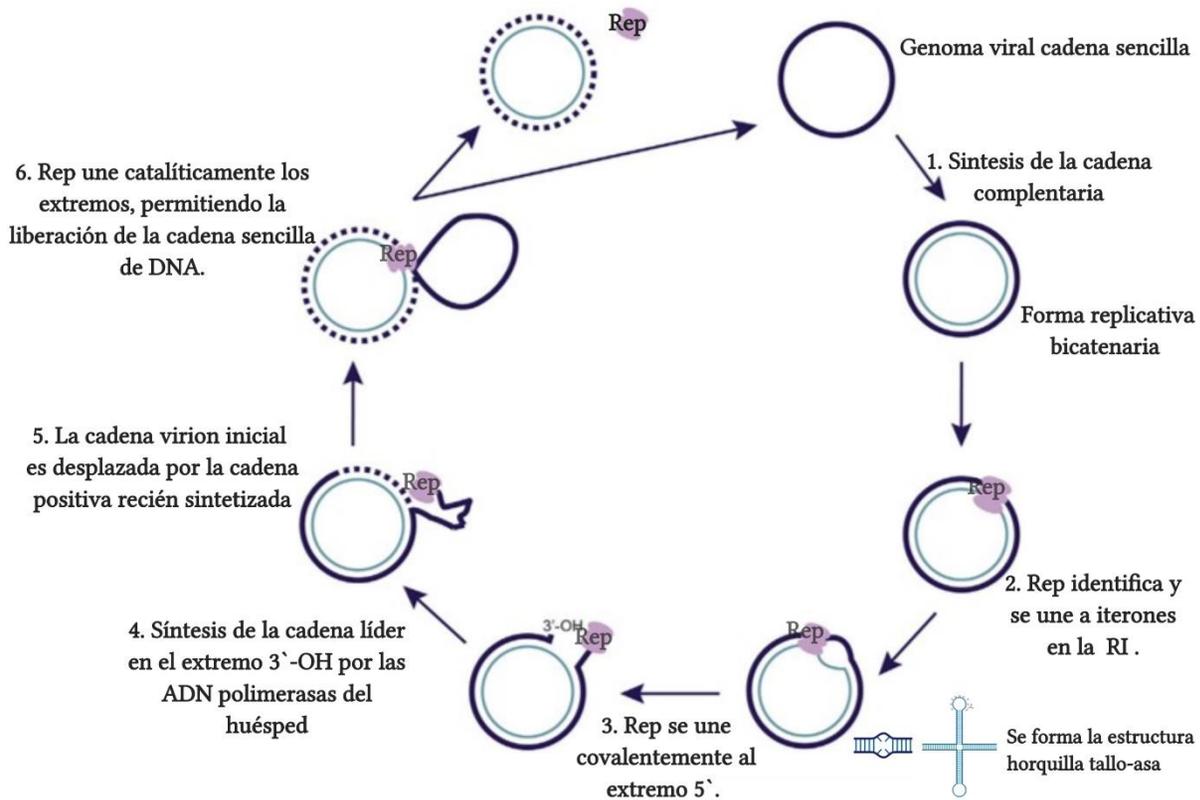


Figura 5. Mecanismo de replicación por círculo rodante. Se representa el genoma viral de cadena sencilla en círculo púrpura grueso. La cadena complementaria se muestra como el círculo azul más delgado. Rep como un círculo púrpura claro. La cadena positiva recién sintetizada en línea púrpura gruesa discontinua. (Adaptado de Zhao *et al.*, 2018). Creado con BioRender.com.

ANTECEDENTES DIRECTOS

Proteínas Iniciadoras de la Replicación (Rep) y los determinantes de especificidad de unión al DNA.

Las proteínas iniciadoras de la replicación por CR, genéricamente denominadas “proteínas Rep”, pertenecen a una Superfamilia muy antigua de Endonucleasas HUH, que se caracterizan por un arreglo definido de tres motivos conservados (Motivos RCR I, II y III) (Ilyina y Koonin 1992; Chandle10 2013) (Figura 6) y son codificadas por una gran variedad de entidades genéticas que se replican por el mecanismo de círculo rodante (RCR), entre los que se encuentran plásmidos de bacterias, arqueas, y eucariontes (ej. algas rojas), así como bacteriófagos y virus de hongos, plantas, y animales (Chandler *et al.*, 2013, Elmer *et al.*, 1988, Londoño *et al.*, 2010, Nash *et al.*, 2011).

La proteína Rep de los geminivirus es una proteína multifuncional (Clerot y Bernardi, 2006; Choudhury *et al.*, 2006; Hanley-Bowdoin *et al.*, 2013; Ruhel and Chakraborty,

2019), que actúa en la RCR como endonucleasa sitio-específica de ssDNA, y tiene actividad de ligasa y como helicasa/ATPasa (Desbiez *et al.*, 1995). Adicionalmente presenta distintos dominios funcionales que le confieren la capacidad de reprogramar del ciclo celular de la planta a través de la interacción con la proteína retinoblastoma, un regulador del ciclo celular (Kong *et al.*, 2000; Arguello-Astorga *et al.*, 2004) y modular las vías de defensa de la planta a través de modificaciones postraduccionales, tal como la fosforilación por cinasas de la planta (Kushwaha, 2017), entre otras actividades. Su función primaria es la de iniciar la replicación viral por el mecanismo de RCR.

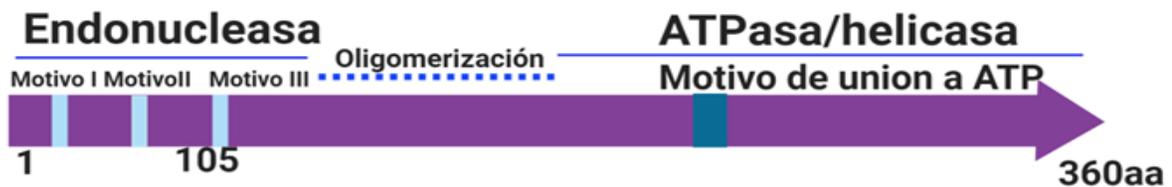


Figura 6. Representación esquemática de la proteína de inicio de la replicación. En el extremo N-terminal se ubica su dominio de endonucleasa, y se resaltan los motivos de unión a DNA conservado en la familia HUH (Motivo I, II y III). Modificada de Rizvi *et al.*, 2015

Como ya se mencionó, la proteína Rep de geminivirus introduce un corte endonucleolítico (“nick”) en la cadena (+) del DNA viral (Fontes *et al.*, 1992, 1994; Laufs *et al.*, 1995; Orozco y Hanley-Bowdoin 1998), en una posición bien definida e invariante dentro de un elemento tallo-asa presente en la región intergénica de esos virus (Figura 7). Para realizar el corte endonucleolítico en el DNA viral, la Rep debe unirse primero a sus sitios de unión de alta afinidad en el intermediario dsDNA del genoma viral, los cuales son elementos repetidos de 5 a 8 pb, organizados en una configuración característica de linaje biogeográfico (ej. virus del Nuevo y Viejo Mundo) y situados en la vecindad del elemento “tallo-asa” conservado (Figura 7). Los “iterones” exhiben una secuencia nucleotídica que varía entre distintos geminivirus, pero que es característica de cada especie o cepa viral, por lo que funcionan como determinantes en *cis* de la especificidad replicativa de los geminivirus (Argüello-Astorga *et al.*, 1994a, 1994b; Rizvi *et al.*, 2015).

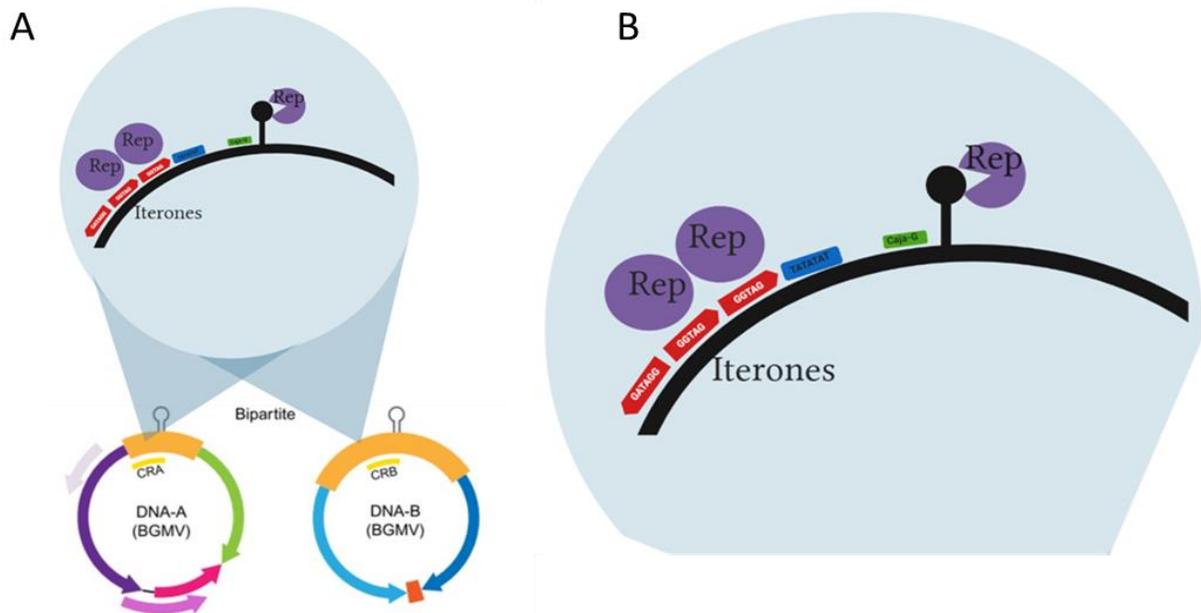


Figura 7. Organización de los sitios de unión (iterones) y de corte (sitio “Nick”) de la proteína Rep en la región común de ambos componentes virales. (A). determinantes en *cis* en la región común de DNA-A y DNA-B. **(B).** Representación esquemática de los elementos presentes en la región común. Las flechas rojas representan los iterones, el recuadro azul, la caja TATA, el recuadro verde la caja G, el círculo morado representa a la proteína Rep. Creado con BioRender.com.

Determinantes en *cis* de la especificidad replicativa de geminivirus.

Muy diversos estudios han demostrado que el ADN-A de un begomovirus puede *trans*-replicar ADN-B heterólogos si éstos poseen iterones similares a los propios, aun cuando los virus sean de linajes diferentes. Este proceso es conocido como pseudo-recombinación. En contraste, existen begomovirus estrechamente relacionados, pero que no son capaces de formar pseudo-recombinantes viables si sus iterones tienen secuencias diferentes. Por ejemplo, dos aislados de *Euphorbia mosaic virus* (EuMV) cuyos 5-6 iterones difieren en un solo nucleótido (GGTGTCC y GGAGTCC, respectivamente), no produjeron pseudo-recombinantes infecciosos en plantas (Gregorio-Jorge *et al.*, 2010). Similar incompatibilidad replicativa se observó entre cepas de ToLCNdV (Chatterji *et al.*, 1999) y entre especies relacionadas como ToMoV de Florida y ToMoTV de Cuba (Ramos *et al.*, 2003). Más recientemente, se demostró que dos cepas de Rhynchosia golden mosaic Sinaloa virus que presentan iterones claramente diferentes, fueron incapaces de formar pseudorecombinantes infecciosos (Avalos-Calleros *et al.*, 2021). En cambio, los reportes de pseudo-recombinación eficiente entre especies virales distintas que poseen iterones similares, son frecuentes (Andrade *et al.*, 2006; Briddon *et al.*, 2011, Chakraborty *et al.*, 2008; Choi *et al.*, 1996,

García-Andrés *et al.*, 2007; Gilbertson *et al.*, 1993; Saunders *et al.*, 2002; 2008; Venkataravanappa *et al.*, 2015; Xu *et al.*, 2019)

Determinantes en *trans* de la especificidad replicativa de begomovirus

La identificación de los iterones en el origen de replicación viral acrecentó el interés para delimitar con precisión los determinantes en *trans* de la especificidad replicativa de los geminivirus, es decir, los residuos aminoacídicos de Rep que interactúan con ellos y determinan, por tanto, su unión específica al ADN de doble cadena (Chatterji *et al.*, 1999; Choi y Stenger, 1995). Experimentos con proteínas híbridas en la década de los 90's mapearon esos determinantes de especificidad dentro de los primeros 116 aminoácidos (Fontes *et al.*, 1992, Choi y Stenger 1996; Jupin *et al.*, 1995). En un estudio teórico, en el que se analizaron *in silico* las proteínas Rep de ~120 geminivirus por medio de un enfoque heurístico-comparativo, se logró delimitar un segmento hipervariable en el extremo N-terminal de Rep, cuya secuencia de aminoácidos correlacionaba significativamente con la secuencia nucleotídica de los iterones cognados. Este "dominio-relacionado al iterón" (IRD, Iteron-Related Domain) se encuentra adyacente al Motivo RCR-I, y se postuló la hipótesis de que residuos particulares del IRD son determinantes de especificidad de Rep (Arguello-Astorga y Ruiz-Medrano, 2001). Esta hipótesis recibió apoyo indirecto con la determinación de la estructura terciaria del dominio catalítico (4-120) de la proteína Rep de TYLCSV, que mostró que el IRD forma parte de una mini-lámina beta que los autores del estudio sugirieron podría estar involucrado en la unión específica al DNA (Campos-Olivas *et al.*, 2002).

La hipótesis del IRD se sometió a prueba experimental por nuestro grupo de trabajo, construyendo ADN-A híbridos a partir del componente genómico A del virus *Tomato mottle virus* (ToMoV), en el que se sustituyó la región intergénica y los primeros 11 codones del gen *AC1* (*Rep*) por el segmento equivalente del ADN-A de los begomovirus PHYVV, ToMoTV y TGMV, que poseen iterones diferentes. Así, los ADN-As híbridos codifican una proteína Rep que es idéntica a la de ToMoV excepto en los primeros 11 residuos (que incluyen al IRD). En experimentos de inoculación de plantas por biobalística, el ADN-A de ToMoV no formó pseudo-recombinantes infecciosos con los ADN-B de PHYVV, ToMoTV y TGMV, como cabía esperar por la diferencia de sus iterones. En cambio, los ADN-A híbridos ToHib-PHYVV, ToHib-ToMoTV, y ToHib-TGMV, produjeron una infección sistémica en las plantas al co-inocularse con el ADN-B del virus cuyo IRD fue fusionado a la proteína Rep de ToMoV (Figura 8), lo cual confirmó la hipótesis de que algunos residuos en el dominio 1-11 son determinantes de especificidad (DEs) replicativa de los virus examinados (Bañuelos-Hernández, 2012).

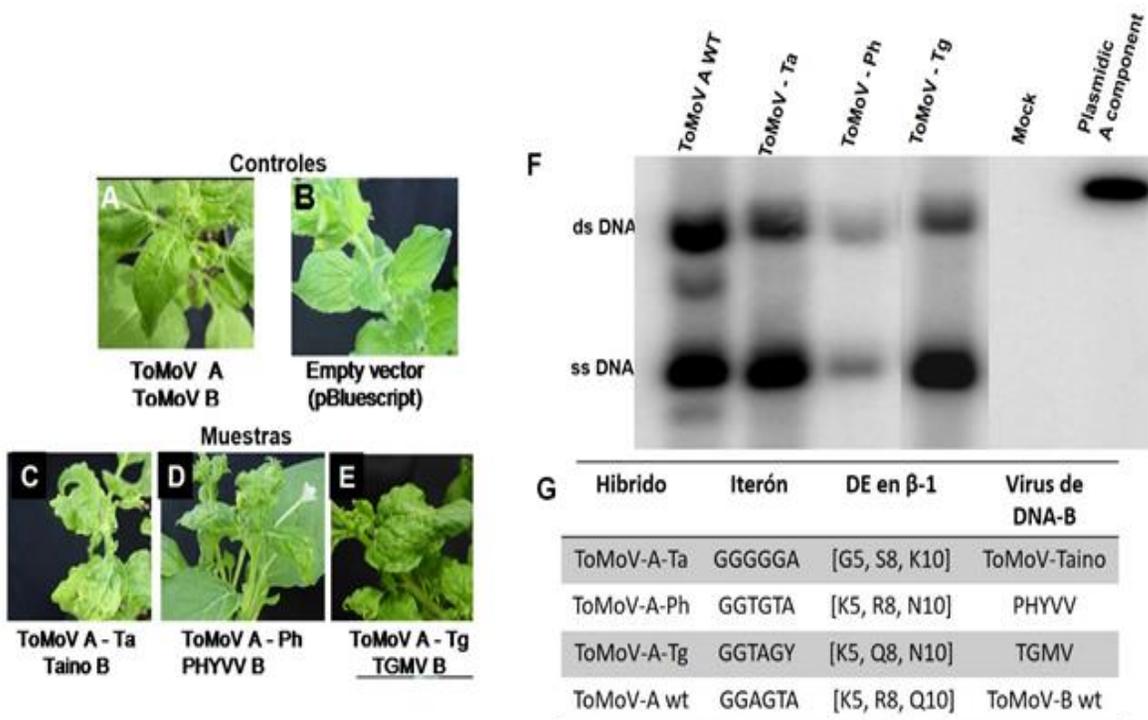


Figura 8. Auto-replicación en células de *N. tabacum* e infección de plantas de tabaco de los componentes genómicos híbridos de ToMoV. (A-E). Plantas de *N. benthamiana* infectadas con componentes silvestres (A) y componentes híbridos y DNA-B correspondiente de las especies virales respectivas (C-E). (F). Autoradiografía de ensayos de autoreplicación de los componentes A híbridos y silvestres en células NT-1. (G). Datos relevantes de los componentes A-híbridos construidos con sus respectivos DEs e Iterones. (Modificado de Bañuelos-Hernández, 2012)

Identificación de potenciales DES de Rep asociados al Motivo RCR I y RCR II.

Londoño *et al.* (2010) llevaron a cabo un análisis comparativo más extenso que incluyó cerca de 180 begomovirus, lo que condujo al descubrimiento de dos SPDs (del inglés "Specificity Determinant") o DEs (en español) adicionales que fueron mapeados a 10 y 12 residuos de distancia, respectivamente, del núcleo HuH del Motivo RCR. Posteriormente se generaron y compararon modelos *in silico* de la estructura tridimensional predicha de algunas proteínas Rep de begomovirus, todas ellas basadas en la estructura 3D de la proteína homóloga de *Tomato yellow leaf curl virus* (TYLCV) (Campos-Olivas *et al.*, 2002). Las proteínas que se compararon usaron iterones diferentes, y los datos estructurales sugirieron que los cinco residuos específicos identificados como potenciales DES en el análisis global de Londoño *et al.*, (residuos en las posiciones 5, 8, 10, 69 y 71 de una proteína begomoviral típica, como la de BGMV) podrían ser, en efecto, los residuos críticos en la interacción de Rep con el ADN. En la figura 9 se presenta un modelo de la estructura terciaria del dominio 1-120 de la proteína Rep de *Sida mosaic Sinaloa virus* (SiMSinV), que reconoce iterones GGGGGA, y se compara en la región del IRD/ beta-5 con el modelo de la proteína homóloga del *Chino del tomate virus* (CdTV), cuyos iterones cognados tienen la

secuencia consenso GGGGTA, que difiere en dos nucleótidos (Mauricio-Castillo *et al.*, 2014).

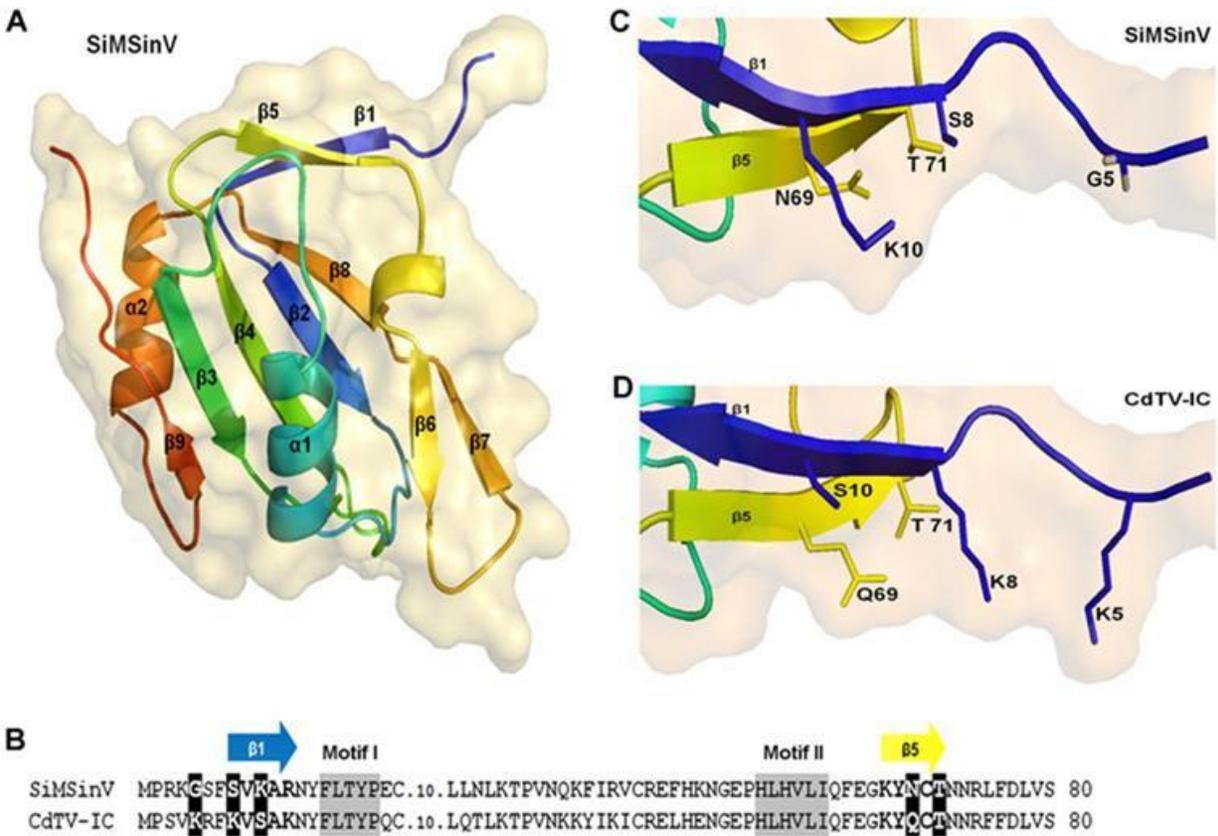


Figura 9- Modelo de la estructura 3D predicha del dominio de unión al ADN de la proteína Rep de SiMSinV. (A) Estructura terciaria del dominio 1-120, usando como templatado la solución NMR del dominio endonucleolítico de la Rep de TYLCSV **(B)** Alineamiento de la region 1-80 de las proteínas Rep de SiMSinV y CdTV. **(C-D)** expansión de la región IRD/ hebra-beta5 de la proteína de SiMSinV (C) y CdTV (D, respectivamente (reproducido de Mauricio-Castillo *et al.*, 2014).

En la Figura 10, se comparan los extremos N-terminales de las proteínas Rep de geminivirus nativos de diferentes continentes, que ilustra la correlación entre los hipotéticos DEs de Rep con la secuencia consenso de sus iterones cognados. Estas variaciones paralelas en las secuencias de macromoléculas diferentes (ADN y proteína) no son fáciles de explicar de manera diferente a la que sugiere el análisis comparado, esto es, que los residuos homólogos a los ubicados en las posiciones 5, 8, 10, 69 y 71 de la proteína Rep de SiMSinV son, muy probablemente, DEs de los iniciadores RCR codificados por los geminivirus (Mauricio-Castillo *et al.*, 2014).

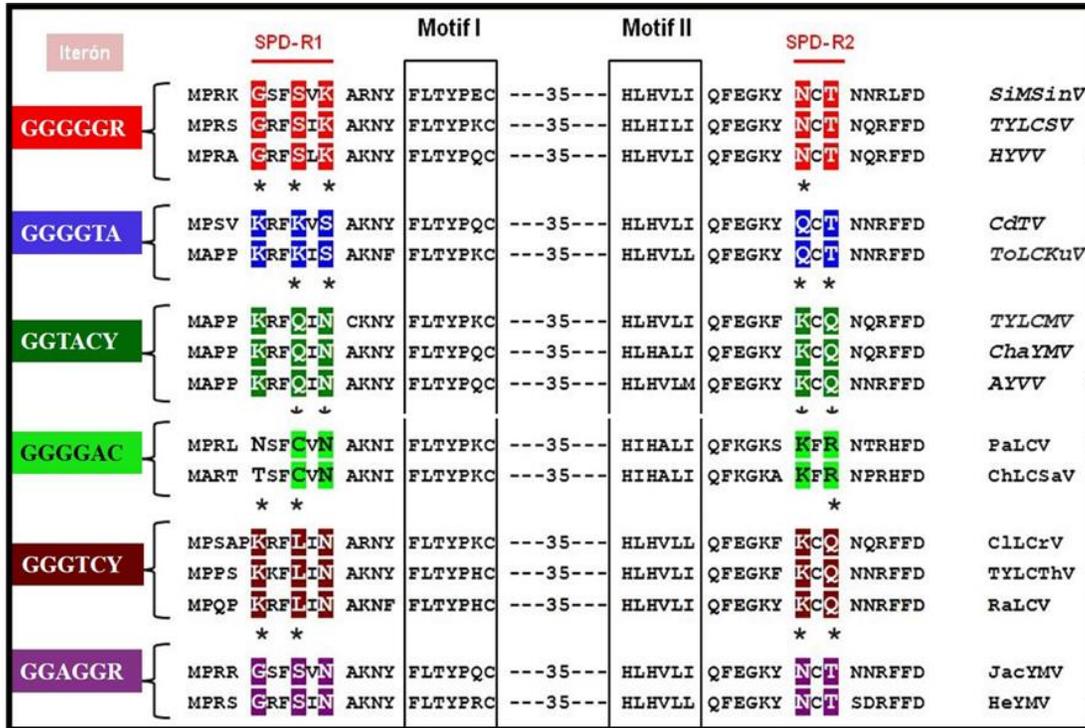


Figura 10- Alineamiento de la región N-terminal de proteínas Rep agrupadas por la similitud de los aminoácidos 5, 8, 10, 69 y 71. La secuencia consenso de los iterones cognados se muestra en el lado izquierdo de la figura. Todos los grupos mostrados incluyen virus nativos de, al menos, dos continentes diferentes.

De todo esto se deduce, naturalmente, **que la combinación específica de los aminoácidos en las posiciones equivalentes a los residuos 5, 8, 10, 69 y 71 de una proteína Rep típica de begomovirus, podrían determinar la unión preferente de Rep a un iteron específico, lo que constituye, por tanto, la hipótesis central del proyecto de investigación cuyo reporte se expone en este trabajo.**

JUSTIFICACIÓN

Al presente solo se ha identificado experimentalmente un determinante de especificidad replicativa de acción en *trans*, esto es, el residuo aminoácido en la posición 10 de la proteína Rep codificada por dos cepas del begomovirus *Tomato leaf curl New Delh virus* (Chatterji *et al.*, 1999). La investigación propuesta aquí proporcionaría información sobre un número muy amplio de determinantes de especificidad replicativa de geminivirus, si se demuestra la validez de la hipótesis central del proyecto.

HIPÓTESIS

- 1) Los residuos aminoácidos de la proteína Rep del geminivirus *Tomato mottle virus* (ToMoV). que determinan su elevada afinidad por los iterones cognados (GGAGTA) se localizan en las posiciones 5, 8, 10, 69 y 71.
- 2) La sustitución selectiva de uno o más de los residuos señalados como potenciales DES en las proteínas Rep de ToMoV, generará proteínas mutantes con una especificidad de unión al ADN alterada y, dentro de ciertos límites, predecible.

OBJETIVOS

Objetivo General

-Someter a prueba experimental la hipótesis de que residuos aa presentes en el llamado IRD (que incluye a la hebra- β 1) y la hebra- β 5 de las proteínas Rep de geminivirus, son críticos para determinar la elevada especificidad de unión a secuencias de ADN definidas en el origen de replicación de esos virus.

- Descifrar, al menos en parte, el código molecular que regula las interacciones específicas de Rep con su origen de replicación cognado.

Objetivos Particulares.

1) Generar mutantes de la proteína Rep del begomovirus ToMoV con sustituciones específicas en los residuos 5, 8, 10, 69 y 71, y determinar por métodos moleculares y ensayos de infección, las secuencias de ADN que dichas proteínas mutadas unen con elevada afinidad.

2) Determinar la capacidad de las proteínas Rep mutantes de promover la replicación de componentes genómicos B de begomovirus con iterones diferentes a los de ToMoV-B, en experimentos de pseudo-recombinación en plantas.

MATERIALES Y MÉTODOS

Material biológico

En este trabajo, el virus del moteado del tomate (ToMoV), proporcionado gentilmente por la Dra. Jane Polston (University of Florida) fue usado como modelo de begomovirus bipartita (Genbank accession: L14460 y L14461)

El componente genómico A de ToMoV fue clonado en el sitio *Apal* del vector pBSKΔXN (un vector pBlueScript-SK del que se eliminó previamente un sitio *Xmnl* de los dos presentes en ese vector, y el único sitio *NgoMIV* localizado en una region funcionalmente irrelevante). El DNA-B se clonó en pBSKΔXE (un derivado de pBlueScript del que se eliminaron los sitios *XbaI* y *EcoRI* del polilinker, proporcionado por la Dra. Nadia Martinez-Marrero). A partir de estas clonas se realizaron las mutaciones que se describen abajo.

La cepa TOP10 de *E. coli* fue utilizada para propagar y almacenar la mayoría de las construcciones virales. Las condiciones de crecimientos fueron en medio Luria- Bertani (NaCl 10g/L, Extracto de levadura 5 g/L y pectona de caseína 10 g/L) por 16 hrs a 37°C en agitación constante a 90 rpm (Sambrook y Russell, 2001). Se adicionaron los antibióticos de selección dependiendo del vector, y a las concentraciones recomendadas. Para el vector pBlueScript SK (-) y sus derivados se seleccionaron con carbenicilina 100 µg/ml.

El vector binario pCAMBIA1300 (donado por el Dr. Bernardo Bañuelos-Hernández) se usó para clonar los hemidímeros de las clonas infectivas de los componentes A y B mutantes, así como el componente silvestre de ToMoV. Las clonas con las construcciones se crecieron y se seleccionaron en medio LB más Kanamicina a 50 µg/ml, cuando están transformadas en *E. coli* TOP10

Para algunos ensayos, fue necesario usar el vector binario pCAMBIA1300 en células de *Agrobacterium tumefaciens* GV3101. En este caso, las bacterias se crecieron en medio SOC (NaCl 0.5 g /L, Extracto de levadura 5 g/L, Triptona 10 g/L, MgCl 10 mM, KCl 0.1 mM y Dextrosa 1 mM) (Sambrook y Russell, 2001) más los siguientes antibióticos: Gentamicina 15 µg/ml, Rifampicina 20 µg/ml y Kanamicina 50µg/ml.

En los ensayos de replicación, se usaron células en suspensión de *Nicotiana tabacum* NT1, crecidas en medio NT1 (sacarosa 15g/L, 10 ml de soluciones Stock: Solución I Nitratos, Solución II Sulfatos, Solución III Quelatos, Solución IV Fosfatos, Solución V Halógenos y ácido di-clorofenoxiacético, 2mg/L-1 pH 5.4). Se incubaron a 25 °C en agitación constante a 90 rpm. Para mantener las células NT1 después de la electroporación el medio NT1 fue suplementado con Manitol a una concentración de 0.4 M (Shimada-Beltran, 2007).

Para los ensayos en discos de hoja y de infección se utilizaron plántulas de 4 semanas de *Nicotiana benthamiana*. incubadas a 25° C en condiciones de 16 hrs de luz/ 8 hrs oscuridad.

Condiciones de OE-PCR

Las condiciones del PCR de extensión traslapada (OE-PCR) (Braman, *et al.*, 2010) para el paso 1 (combinaciones FNgoMIV/M13R y FM13/RNgoMIV (Tabla S1 de anexos) fueron las siguientes: se usó MyTaq™ Red DNA polymerase (Bioline, Meridian Life Science, Memphis, USA). El buffer contiene 50mM de dNTPs, 15 mM de MgCl₂,

agentes estabilizadores (concentraciones optimas según el fabricante) y las condiciones del termociclador: 95°C 1 min (1 ciclo); 95°C 15 seg, 55°C 11 seg, 72°C 100 seg por 25 ciclos y 4° ∞. Ambos fragmentos fueron agregados en una relación 1:1 (Concentración final 50 ng); el segundo paso en el que se utilizaron los iniciadores FM13 y M13R en las siguientes condiciones 98°C 2 min, 98°C 30 seg, 55°C 30 seg, 72°C 30 seg por 20 ciclos 72°C 5 min y 4° ∞, se usó la *Taq* polimerasa platinum (Invitrogen) según las especificaciones del proveedor

Generación de casetes sintéticos de la región IRD del gen *Rep*.

Los casetes sintéticos de doble cadena se generaron a partir de oligonucleótidos de cadena sencilla de 33 o 37 pb (Tabla S2), diseñados con cambios en los codones que corresponden a los aminoácidos de interés (5, 8 y 10). Cada pareja de oligonucleótidos de cadena sencilla se hibridó, usando una concentración de 1µM de cada primer más buffer de hibridación 1X (Tris-HCl 200mM pH 7.5, NaCl 1.5M) y se incubaron 5 minutos a 95°C; se dejó que la temperatura descendiera gradualmente hasta llegar a temperatura ambiente; Los casetes de dsDNA se fosforilaron usando la T4 polinucleótido cinasa (New England, Biolabs) de acuerdo con las instrucciones del proveedor.

Condiciones para PCR mutagénico (Q5 Site-Directed Mutagenesis Kit)

Se diseñaron los oligonucleótidos mutagénicos (Tabla S1) para los cambios en los codones que corresponden a los aminoácidos 69 y 71 de la proteína *Rep*. En todos los casos se utilizó el mismo el iniciador “reverso” (RniPciI). Las mutaciones se realizaron por PCR mutagénico sitio dirigida usando Q5 Site-Directed Mutagenesis Kit New England (ver anexos); el DNA usado como templado fueron los componentes A mutantes en los codones 5, 8 y 10, previamente generados, que corresponden a cada combinación específica examinada. Los oligonucleótidos (forward y reverso) se prepararon a una concentración stock de 10 pM cada uno, posteriormente se fosforilaron usando la T4 polinucleotide Kinase (según indicaciones del proveedor), se incubo a 37°C por una hora, después se inactivo la enzima agregando 1 volumen de Cloroformo-Alcohol Isoamilico 24:1 y se mezcló usando vortex por 1 minuto y se centrifugó a 13000 rpm por 10 minutos Se tomó la fase superior y se colocó en un tubo nuevo; esa fase es la que contiene los iniciadores fosforilados, los cuales se usaron en la PCR mutagénico (Sambrook, 2001). La mezcla de reacción para el PCR fue la siguiente: Buffer HF 1X, dNTP`'s 200 µM, oligo Forward 0.5 µM y Reverso 0.5 µM, DNA templado 2ng/0.5 µL, Q5® *High-Fidelity DNA Polymerase* 0.02U/µL Las condiciones de la PCR fueron las siguientes: 98°C 1 minuto; 98°C 30 seg, 68.5°C 30 seg, 72°C 1.45 minutos por 35 ciclos; 72°C 10 min y 4° ∞. Los productos de PCR fueron tratados con 10 U de *DpnI*, se incubo 3 horas a 25 °C; posteriormente se inactivó la enzima a 65°C por 10 minutos, para eliminar el DNA templado proveniente de la bacteria. Por último, los productos de cada reacción tratados con la endonucleasa se ligaron usando Ligasa T4 (NewEngland), se incubo a 25°C por 2 h, esta ligación fue transformada en *E. coli* TOP 10 y las clonas fueron seleccionadas por medio de un corte *PciI* que introduce el oligonucleótido reverso, y enviadas a secuenciar para confirmar que las mutaciones fueron incorporadas.

Generación de casetes sintéticos con secuencias de iterones diferentes a los de ToMoV por PCR de fusión.

Las secuencias nativas de los iterones en la región río arriba del gen *Rep* en el componente genómico A, o río arriba del gen *BC1* en el componente B, se reemplazaron por casetes sintéticos de doble cadena de aproximadamente 150 pb generados por PCR de fusión.

Para el diseño de los casetes sintéticos se usó como plantilla la secuencia nativa de un fragmento *NgoMIV/ PmlI* del componente A de ToMoV, donde se ubican los iterones. Se obtuvieron un total de 8 casetes sintéticos cuya secuencia de iterones corresponde a cada DNA-A mutante con DEs distintos. Los segmentos sintéticos se obtuvieron por el traslape en un segmento de 37 nt de mega-oligonucleótidos de 94 nt junto con su mega-oligo complementario para extender los ADNdc a ~150 p. Los mega-oligonucleótidos se enlistan en la Tabla S3 de Anexos.

Las condiciones de la PCR de Fusión fueron: 98°C 2 min, 98°C 30 seg 55°C 30 seg, 72°C 30 seg por 20 ciclos 72°C 5 min y 4° ∞. Cada producto de PCR fue clonado en el vector pGEM-T Easy (PROMEGA), y se verificó por secuenciación.

Construcción de componentes A doblemente mutados (DEs e Iterones)

El origen de replicación del componente A mutante se digirió con *PmlI* y *NgoMIV* por una hora a 37 °C (fragmentos: 5416 pb y ~135pb); se purificó la banda de 5416 pb usando el Kit Wizard (Promega). Se trató con las enzimas *PmlI* y *NgoMIV* al pGEM-T Easy que contenía el casete sintético correspondiente, se incubó a 37°C por una hora, posteriormente los fragmentos (137 pb y ~3000 pb) se separaron en un gel de agarosa al 1.5% y se purificó del fragmento de 137 pb con el Kit Wizard (Promega). Ambos fragmentos fueron ligados (relación 1:3) y se transformó en *E. coli* TOP 10; las clonas resultantes se enviaron a secuenciar para confirmar la substitución correcta de los iterones.

Construcción de Componentes B mutantes.

Se trató el componente B silvestre de ToMoV clonado en el vector pBlue-Script con las enzimas *EcoRI/PmlI* que generaron fragmentos de 5171pb y 338pb, respectivamente), se incubó una hora a 37°C posteriormente se separaron los fragmentos en un gel de agarosa al 1% y se purificó la banda de 5171pb usando el Kit Wizard (Promega); este fragmento se utilizó para insertar los diferentes fragmentos sintéticos con iterones diferentes al silvestre. Se trató con las enzimas *PmlI* y *EcoRI* al pGEM-Easy que contiene el casete sintético, se incubó a 37°C por una hora, posteriormente los fragmentos (137 pb y 3000 pb) se separaron en un gel de agarosa al 1.5% y se purificó el fragmento de 137pb con el Kit Wizard (Promega). Ambos fragmentos purificados se ligaron en una relación 1:3 y se transformaron en *E. coli* TOP 10. Las clonas resultantes se analizaron con el corte *EcoRI/PmlI* (fragmentos esperados: 5171pb y 137 pb) y se enviaron a secuenciar para confirmar el reemplazo. Los componentes B resultantes se nombraron de acuerdo con el virus que presenta iterones como los sintetizados.

Construcción de clonas infectivas de los componentes genómicos A y B mutantes.

Las clonas infectivas están formadas por ~1.5 monómeros de DNA-A o B (hemidímeros) de los virus clonados en un vector apropiado (Kim *et al.*, 2007; Peyret y Lomonosoff 2015; Porta y Lomonosoff, 2002). Para realizar las construcciones con el DNA-A, un fragmento de 1.5 Kb de éste fue amplificado por medio de PCR, usando los primers FTH1K/M13R, y el producto (que contiene 72 pb del gen *CP*, 308 pb que corresponden a la región intergénica, y 837 pb del gen *Rep*) fue clonado en el vector pGEM-Easy y enviado a secuenciar. Posteriormente se subclonó un fragmento *HindIII/EcoRI* (1.5 Kb) del DNA-A en pCAMBIA1300. A la clona resultante se le inserto un monómero del DNA-A en un sitio *Apal*. Se realizó un primer análisis mediante enzimas de restricción para verificar la orientación y posteriormente se confirmaron por secuenciación. Los hemidímeros resultantes se transformaron por electroporación en *A. tumefaciens* GV3101 y se seleccionaron mediante resistencia a Km (50 µg/ml), Gm (40 µg/ml) y Rif (20µg/ml); las clonas resultantes se verificaron mediante PCR en colonia.

Transformación de protoplastos de células NT1 de *Nicotiana tabacum*

Los componentes A mutantes en los residuos 5, 8, 10, y los mutantes en los residuos 69 y 71, fueron examinados en ensayos de auto-replicación en protoplastos de *N. tabacum* NT1. Las células en suspensión de 4 días fueron recolectadas por centrifugación a 2000 rpm por dos minutos y lavadas con manitol al 0.4 M; posteriormente se agregó solución enzimática (manitol 0.4M, MES 20 mM, celulasa 1% y pectoliasa al 0.1%) y se incubó por 40 min a 90 rpm a 23 °C. Las células fueron colectadas por centrifugación a 1600 rpm por 2 min y se lavaron dos veces con solución de manitol al 0.4M, y dos con buffer de electroporación (NaCl 0.8% KCl 0.2%, NaK₂PO₄ 0.2%, Na₂HPO₄ 0.11%, Manitol 0.4M pH 6.5) (Shimada-Beltran 2007). Las células fueron resuspendidas en 2 volúmenes de buffer de electroporación. Se prepararon alícuotas de 300 µL y se colocaron en celdas de electroporación (BIORAD) junto con 10 µg de DNA-A previamente digerido. Las condiciones del pulso fueron: 250 V, 500 µF. Las células electroporadas fueron recuperadas en medio NT (ver arriba) suplementado con manitol al 0.4M e incubadas por 48 hrs a 23 °C en oscuridad. Posterior a la incubación se procedió a la extracción de DNA total y su digestión con *Apal* y *Dpnl*, como se describe más adelante.

Southern blot de ensayos de replicación transitoria en protoplastos

Los protoplastos fueron recolectados en tubos de 1.5 ml por centrifugación a 2500 rpm, el paquete celular fue sumergido en nitrógeno líquido y macerado, se agregaron 480 µl de buffer de extracción de plantas (EDTA 50 mM, Tris-HCL 100 mM) y 37.5 µl de SDS al 20 % y se incubaron a 65 C por 10 minutos, después se agregaron 94 µl de Acetato de potasio 5 M y se incubó en hielo por 5 minutos, posteriormente se centrifugaron a 13 000 rpm por 10 minutos y el sobrenadante se transfirió a un tubo nuevo de 1.5 ml y se agregó un volumen de Fenol-Cloroformo y se agitó en vortex por un minuto. Se centrifugó por 5 minutos a 13 rpm y la fase acuosa se colocó en un tubo nuevo, se agregó 600 µl de isopropanol frío y se incubó a -20 C por 20 minutos, después se centrifugó a 13 000 rpm por 20 minutos. Se eliminó el sobrenadante y la pastilla se lavó con 500 µl de etanol al 70% centrifugando por 5 minutos. Por último, se

decantó y se dejó secar la pastilla a temperatura ambiente y se resuspendió en 35 μ l de agua MiliQ (Dellaporta, *et al.*, 1983).

Posteriormente, se trató con las enzimas *DpnI* y *Apal* a 37 C toda la noche. El DNA fue sometido a electroforesis en gel de agarosa al 1% y transferido a membranas de nylon Hybond+ (Amersham) usando los protocolos estándar. Se usó como sonda un fragmento específico del componente A de ToMoV (fragmento *Apal/SacI* de 600 pb, correspondiente a un segmento del gen rep.) La sonda fue marcada con biotina, usando “Biotin 3’DNA Labeling Kit” (ThermoScientific) de acuerdo con las condiciones sugeridas por el proveedor, y detectada por quimioluminiscencia usando “Chemiluminescent Nucleic Acid Detection Kit” (ThermoScientific). Para la detección del componente B se usó como sonda un fragmento *Apal/BamHI* de 600 pb, el cual se trató exactamente igual que la sonda para el componente A.

Ensayos de autoreplicación en discos de hojas (Leaf disk assays)

Las cepas de *Agrobacterium* que contienen las clonas infectivas de los componentes A mutantes y silvestre, fueron crecidas en medio SOC más Km (50 μ g/ml), Gm (40 μ g/ml) y Rif (25 μ g/ml), por 18 horas. Las bacterias fueron recolectadas por centrifugación a 4000 rpm por 10 minutos, después se realizaron dos lavados con buffer de agroinfiltración (Acetosiringona 10 μ M, MgCl 10 mM, MES 10mM pH 5.7) para eliminar los desechos del medio. Posteriormente se ajustaron a una $DO_{600}=1$. Se agroinfiltraron 1ml de la suspensión bacteriana en cada hoja de cada planta (Kim *et al.*, 2007). Se incubaron por 72 horas en cámara ambiental a 23°C en condiciones estáticas (Xu *et al.*, 2018; Elmer *et al.*, 1988) A continuación, las hojas se recolectaron para extraer DNA y realizar ensayos de Southern blot para detectar si hubo replicación del componente A.

Ensayos de infección de *N. benthamiana*

Los ensayos de infección se realizaron en plantas *N. benthamiana* de 4 semanas de edad. El número de muestras fue de 9 plantas para el control positivo y las muestras de las diferentes combinaciones pareadas DNA-A y DNA-B mutantes, mientras que para los controles negativos (mock y sin tratamiento) fue de 7 muestras. Las réplicas técnicas fueron realizadas por triplicado. Las clonas infectivas de los DNA-A y B silvestres, y las 5 mutantes (Tipo ACMV, BCTV, OkYMMV, PHYVV y TGMV), transformadas en *A. tumefaciens* GV3101, fueron crecidas en medio SOC más Kanamicina (50 μ g/ml) y Rifampicina (20 μ g/ml) a 28C por 16 h, después las bacterias se recolectaron por centrifugación a 4000 rpm por 10 minutos y se les realizaron 2 lavados con la solución de Agroinfiltración (Acetosiringona 10 μ M, MgCl 10 mM, MES10mM pH 5.7). Posteriormente se ajustaron a una $D.O.600 =1$. Se realizó una mezcla de las suspensiones bacteriana en una relación 1:1. Se agroinfiltraron 500 μ l de la mezcla en 3 hojas de cada planta. Se incubaron en una cámara ambiental con las siguientes condiciones: ciclo luz/oscuridad 12 h, humedad 25% y temperatura a 25 C (Elmer, *et al.*, 1988) A). Se tomaron muestras de hojas nuevas a los 10, 14 y 21 días post- infección (dpi), para la detección de los componentes A y B de las mutantes y el virus silvestre. A las hojas nuevas, se les extrajo ADN total, como se menciona arriba, 100 ng de éste se usó como templado para la detección DNA *de novo* por medio de PCR. Se usaron las siguientes combinaciones de iniciadores para amplificar el segmento superior del DNA-A DGR-For / YMAC-Rev (Fragmento de 1404pb); DNA-A

segmento inferior YMAC-For / DGR-For; Segmento superior de DNA-B BC-290-For/BV-310-Rev y segmento inferior DNA-B segmento inferior BV-310-For/ BC-290-Rev. Las condiciones de PCR fueron: 95°C 2 min, 95°C 1 min, 55°C 1 min, 72°C 2 min por 25 ciclos 72°C 5 min y 4°C. Los productos fueron visualizados en un gel de Agarosa al 1%. Los fragmentos que corresponden a los segmentos superiores fueron clonados en el vector pGEM-T Easy (PROMEGA) y enviados a secuenciar.

Análisis de secuencias.

Los mapas del genoma del virus del moteado del tomate (ToMoV), sus mutantes y fragmentos clonados en los vectores pCAMBIA1300, pGEM-T Easy y pBSK(-)SK fueron construidos usando SnapGene® o SnapGene Viewer (<http://www.snapgene.com>; GSL Biotech, Chicago, IL, EE. UU.) (Sturmberger et al., 2016). La comparación de genomas se realizó con BLASTn (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Los alineamientos entre secuencias se realizaron en CLUSTAL OMEGA (<https://www.ebi.ac.uk/Tools/msa/clustalo/>)(Li 2003). Se utilizaron ajustes predeterminados en todos los softwares.

RESULTADOS

Tomando en consideración el gran número de modificaciones que fue necesario realizar al genoma del virus Tomato mottle virus (ToMoV), consideramos importante ofrecer una visión panorámica de los cambios realizados, intentando ser claros para la comprensión global del proyecto de investigación que llevamos a cabo.

El proceso de generación de las mutantes de ToMoV siguió un orden muy preciso, por lo que omitiremos mencionar los aspectos muy técnicos que interrumpirían la fluidez y claridad de la descripción. Esos detalles técnicos se describen en apartados fácilmente localizables en los apartados creados para ese fin.

A grandes trazos, se llevaron a cabo las siguientes etapas de modificaciones moleculares (mutaciones simples o compuestas):

1. Generación de mutaciones en los **codones 5, 8 y 10** del gen Rep de ToMoV, que resultaron en la sustitución precisa de los residuos aminoácidos originales por otros que se seleccionaron a partir de un Modelo Predictivo General (ver más adelante). Estas mutaciones se realizaron por la técnica de “sustitución por cartuchos sintéticos”.
2. Una vez mutados los codones 5, 8 y 10, los ADN-As mutantes correspondientes se utilizaron como plantillas para generar el **segundo grupo de mutaciones** (codones **69 y 71**) en todos los casos en que los aminoácidos correspondientes fueran diferentes a los de la proteína Rep de ToMoV.
3. Una vez realizadas las mutaciones en el gen Rep de ToMoV, se procedió a reemplazar los cuatro iterones nativos (GGAGTA) de ese virus, por conjuntos alternativos de cuatro iterones, que corresponden al sitio de unión predicho por el Modelo Predictivo para la proteína Rep codificada por cada DNA-A mutante. En otras palabras, los iterones de ToMoV fueron sustituidos por los elementos repetidos que el modelo predice serán unidos eficientemente por la proteína Rep mutante correspondiente.

Tomados en conjunto, el resultado de las operaciones realizadas fue la generación de 8 mutantes globales del DNA-A de ToMoV, los cuales poseen orígenes de replicación diferentes, con iterones que presumiblemente serán reconocidos por la proteína Rep mutada en los aminoácidos 5,8,10, 69 y 71 según una combinación bien definida, que es similar a la de proteínas Rep de geminivirus silvestres con iterones cognados como los predichos (Figura 11).

De acuerdo con la hipótesis central del proyecto, esos componentes genómicos derivados de ToMoV-A serán capaces de autorreplicarse en células vegetales, y de trans-replicar DNA-Bs derivados del componente genómico B de ToMoV mediante la subclonación (“transplante”) del origen de replicación del DNA-A mutante correspondiente. Por lo tanto, los componentes genómicos mutantes (A+B) deberían también ser capaces de producir una infección sistémica en plantas de *N. benthamiana*.

En el período del proyecto solo se lograron generar y verificar 5 mutantes del DNA-B de ToMoV con la región intergénica intercambiada, por lo que los experimentos de infección se restringieron a esos 5 virus mutantes y al virus ToMoV silvestre.

En lo que sigue procederemos a describir en detalle cada fase del proceso de generación de los virus mutantes. Los resultados de los experimentos realizados con los cinco virus mutantes completos (DNA-A + DNA-B) de ToMoV que fueron caracterizados, son descritos en la parte final de presente trabajo.

Selección de determinantes de especificidad en *cis* y en *trans* que se examinaron en el estudio.

La primera actividad para desarrollar el presente proyecto fue definir las combinaciones de los presuntos pares cognados Iterón-Rep que fueran las más apropiadas para examinar la hipótesis central del estudio. Esta elección es crucial, ya que existen decenas de geminivirus con iterones diferentes y que codifican proteínas Rep divergentes, pero algunos de ellos podrían ser mejores que otros para verificar las hipótesis de trabajo. Por lo tanto, tomamos en cuenta varios criterios para que la elección fuese apropiada para el estudio. La primera fue definir el nivel de “robustez” relativa de las correlaciones Rep-iterón que se han descrito en la literatura (Arguello-Astorga y Ruiz Medrano 2001; Londoño *et al.*, 2010; Mauricio-Castillo *et al.*, 2014), Una correlación Rep-Iterón se considera “altamente robusta” cuando se encuentra en un grupo comparativamente grande (>8) de virus no relacionados, y con el mínimo número aparente de excepciones a la regla definida (idealmente ninguna). Un segundo criterio es la preservación de la correlación entre virus pertenecientes a distintos linajes filogeográficos (ej., de diferentes continentes), lo que da mayor peso a las correlaciones. Un tercer criterio es que los virus elegidos como referencia sean especies lejanamente relacionadas al virus modelo, ToMoV, y que sus iterones y sus presuntos DES presenten diferencias desde muy pequeñas (una base del iterón, un aminoácido de Rep) a muy grandes (ej., 4 bases en los iterones, 4 residuos aminoácidos diferenciales en Rep), con los elementos correspondientes de ToMoV, lo que haría factible una valoración más precisa del papel de cada potencial DEs, en caso de discrepancias con la hipótesis central.

En base a esos criterios elegimos ocho proteínas Rep con combinaciones características en los aminoácidos (5, 8, 10) / (69, 71) y cuyo presunto Iterón homólogo (cognado) fuese muy conservado, y claramente distinto al de otros grupos de proteínas Rep. Entre las proteínas seleccionadas incluimos proteínas codificadas por geminivirus de distintos continentes e incluso de diferentes géneros (ej. begomovirus y curtovirus). Los 8 pares cognados Rep-iterón, que seleccionamos después de una revisión minuciosa, satisfacen los requisitos indicados, y se resumen en el Modelo Predictivo ilustrado en la Figura 11. Todas las mutaciones realizadas experimentalmente en el componente genómico A de ToMoV se basaron en ese Modelo Predictivo, y los resultados experimentales obtenidos han de ser examinados a la luz de dicho modelo.

[K5, R8, Q10] [Q69, T71]	GGAGTA	ToMoV
	Iteron predicho	Geminivirus
β-1 β-5		
[K5, R8, N10] [Q69, T71]	GG T GTA	PHYVV
[K5, K8, S10] [Q69, T71]	GG G GTA	CdTV
[K5, K8, Q10] [V69, T71]	GGAG A C	ACMV
[G5, S8, K10] [N69, T71]	GGGG A	SiMSinV
[K5, L8, N10] [K69, Q71]	GG G TCT	TYLCThV
[K5, Q8, N10] [K69, Q71]	GG T ACC	OkYMMV
[K5, Q8, N10] [C69, Q71]	GG T AGY	TGMV
[N5, Q8, K10] [Q69, R71]	GG G TGC	BCTV

Figura 11- Modelo Predictivo adoptado en el presente estudio, que ilustra las diferencias en los potenciales DEs de las proteínas Rep (residuos 5, 8,10, 69 y 71) que correlacionan con un iterón definido en múltiples virus de linajes diversos. En letras azules se muestran los residuos aminoácidos de la proteína Rep de ToMoV, y en letras negras la secuencia del iterón del mismo virus. En letras rojas se denotan los residuos aminoácidos de Rep y los nucleótidos del iterón que difieren de los presentes en los elementos equivalentes de ToMoV. Los acrónimos en la columna derecha corresponden al virus “tipo” que ejemplifica un geminivirus que presenta el Iterón cognado (predicho) y la combinación de residuos [5,8,10/ 69,71] correlativa.

Modificación del componente A de ToMoV para introducir un sitio de corte al inicio del gen Rep.

El componente genómico A de ToMoV (2601 pb) se clonó en el sitio *Apal* del vector modificado pBSKΔXN, el cual deriva del plásmido pBlueScript-SK del cual se eliminaron previamente los sitios de corte para las endonucleasas *NgoMIV* y uno de los dos sitios *XmnI* nativos (ver anexos). La nueva clona fue nombrada pBSKΔXN-ToMoVA. Esta clona posteriormente fue modificada por mutagénesis dirigida al inicio de la secuencia del gen *Rep*, introduciendo un sitio *NgoMIV* en la región de los codones 2 y 3. El nuevo sitio de corte fue generado por extensión de traslape de PCR (OE-PCR) (Braman *et al.*, 2010) usando los oligonucleótidos mutagénicos (FNgoMIV/RNgoMIV) y los iniciadores universales M13F y M13R. El codón 2 original CCC (Prolina) se cambió por CCA, una mutación silenciosa. Con la mutación introducida, el codón 3 CCA (Prolina) se cambió por GCA, generando de ese modo un cambio del aminoácido codificado, Prolina por Alanina (A3P). En la primera ronda de PCR se utilizaron las combinaciones de iniciadores FNgoMIV/RM13 y FM13/RNgoMIV; los productos de PCR fueron de 1855pb y 1021 pb respectivamente. Posteriormente ambos fragmentos se utilizaron como templado para la segunda PCR donde se utilizaron los iniciadores M13F y M13 R, cuyo amplicón fue de 2832 pb. La introducción del sitio *NgoMIV* se verificó digiriendo el producto de PCR extenso con la enzima de interés (fragmentos 1821pb y 1003pb) (Figura 12 B). Por último, el producto de PCR se digirió con *Apal*, para clonarse en un sitio compatible en el vector pBSKΔNX. Las clonas resultantes se analizaron primero por patrón de digestión con *Apal/NgoMIV* (2960, 1764, 837 pb) y

ApaI/XmnI (2960, 1802, 800 pb) (Figura 12B). La construcción resultante se verificó por secuenciación y se le denominó pBSK Δ XN-ToMoVA^{NgoMIV}. (Figura 12). Este plásmido se utilizó como base para realizar todas las modificaciones en el gen *Rep*, a fin de obtener proteínas Rep con distintos determinantes de especificidad (DEs) de unión al DNA.

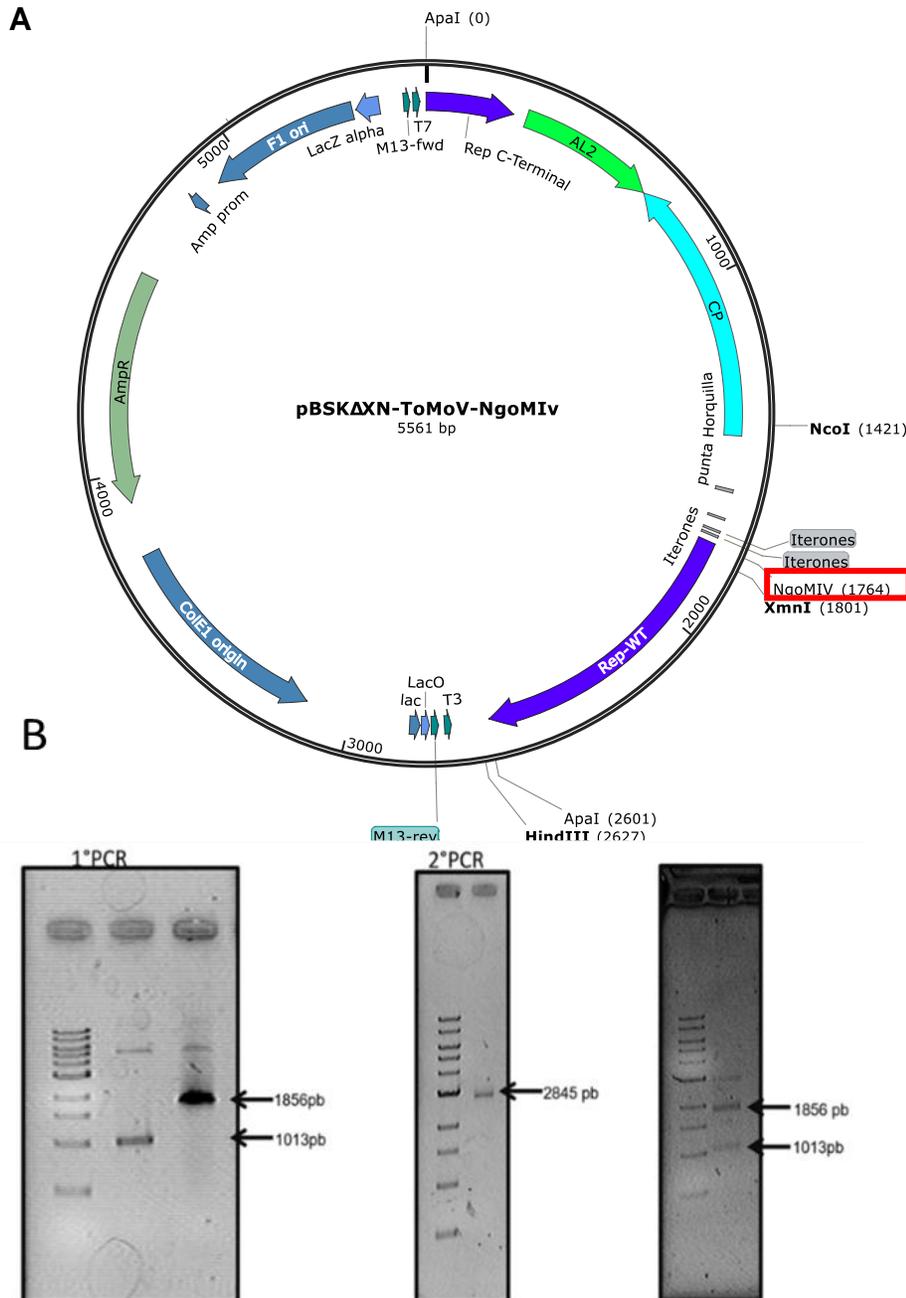


Figura 12. Introducción del sitio *NgoMIV* en la secuencia codificante del gen *rep* en el componente A de ToMoV. (A). Mapa de la clona pBS-ToMoVA-NgoMIV, que incluye un monómero del DNA-A de ToMoV con un sitio único *NgoMIV* introducido al inicio del gen *Rep*, para facilitar el reemplazo del segmento que incluye a los codones 3

a 11. El sitio *NgoM* IV se resalta con un rectángulo rojo. **(B)** Gel 1 Primera ronda de PCR en la línea 1. MPM1kb DNA Ladder (New England Lab), Carril 2, PCR usando oligos M13F/R*NgoM*IV, Carril 3, PCR con oligos F*NgoM*IV/RM13. Gel 2. Segunda ronda de PCR con oligos universales M13F/M13R. carril 1 MPM (NEB). carril 2 producto de PCR de 2845 pb. Gel 3. Carril 1 MPM (NEB), línea 2, Producto de PCR con oligos universales tratado con la enzima *NgoM*IV.

Generación de variantes de ToMoV-A con mutaciones en los codones 5, 8, y 10, del gen Rep.

En la primera etapa del proyecto se generaron mutantes de ToMoV que codifican ocho proteínas Rep diferentes, que presentan combinaciones de potenciales determinantes de especificidad (DES) en el extremo N-terminal, que son distintos a la proteína silvestre, tales como los que presenta el virus del mosaico africano de la yuca (ACMV), el virus del mosaico dorado del tomate (TGMV) y otros geminivirus descritos en la Figura 11.

Las ocho versiones mutantes del gen *Rep* de ToMoV A se generaron en dos pasos para simplificar el proceso; en una primera etapa se cambiaron por substitución de cartuchos o casetes de DNA sintéticos, y en un paso posterior se modificaron los codones que corresponden a los aminoácidos 69 y 71.

El esquema ilustrado en la Figura 13 ilustra la estrategia general utilizada para el reemplazo del casete que comprende los codones 5, 8 y 10, localizados entre los sitios *NgoM*IV y *Xmn*I, lo que hizo factible y técnicamente sencilla la substitución de ese segmento del gen Rep nativo de ToMoV por la región equivalente (“mutante”) de las 8 variantes elegidas-

Se diseñaron 8 casetes sintéticos de 37 pb cada uno, que corresponden a una combinación específica de los DES 5, 8 y 10, como se muestran en detalle en la Tabla S2 de anexos, y en base al Modelo Predictivo ilustrado previamente. Los casetes sintéticos se obtuvieron mediante la hibridación de oligonucleótidos de cadena sencilla de 37 y 33 nt, respectivamente, complementarios en 33 posiciones (ver Materiales y métodos); al formar la doble cadena se generan cartuchos con extremos compatibles con los dejados por las enzimas *Xmn*I (extremo romo) y *NgoM*IV (extremo cohesivo) en la secuencia de ToMoVA-*NgoM*IV (Figura 13).

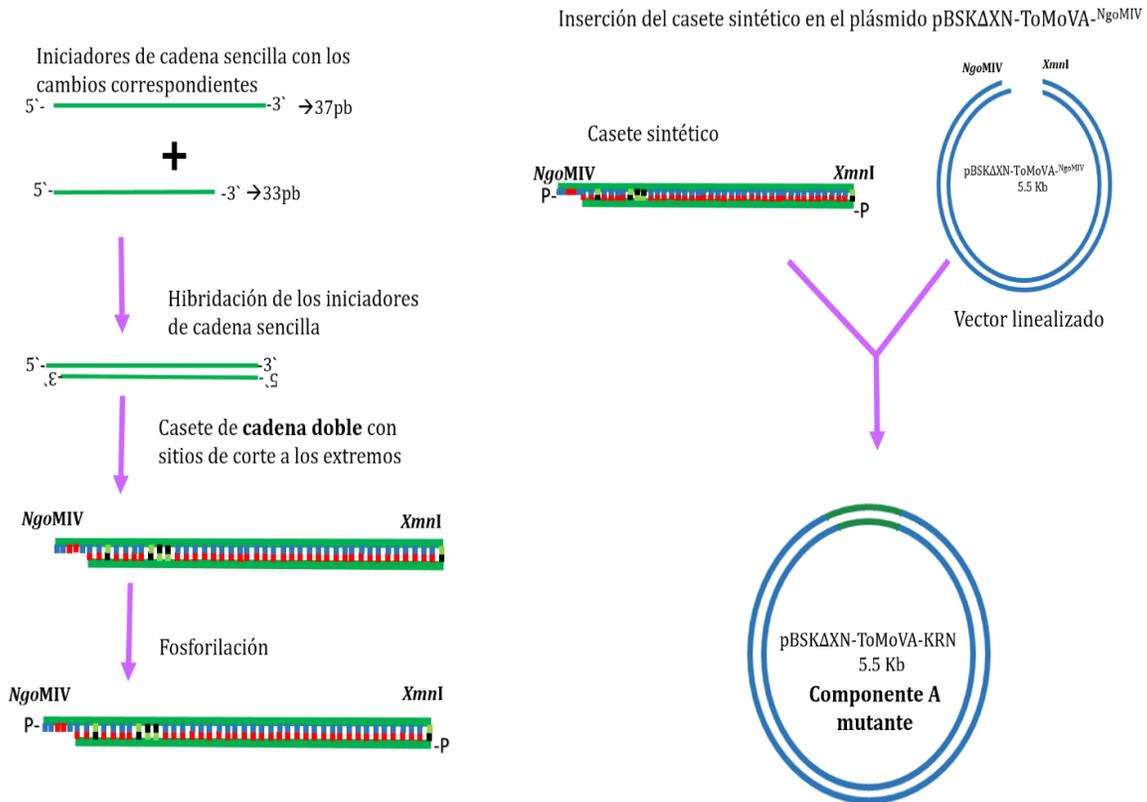


Figura 13. Estrategia para el reemplazo de 37 pb del inicio del gen *Rep* mediante casetes sintéticos mutantes. Oligonucleótidos de cadena sencilla fueron hibridados con su iniciador complementario, para generar los casetes de doble cadena con extremos compatibles *NgoMIV* y *XmnI*. Por último, los casetes se fosforilaron y ligaron con el plásmido pBSKΔXN ToMoVA-^{NgoMIV} cortado con las enzimas mencionadas.

El vector con el componente A-^{NgoMIV} se digirió con ambas enzimas y se purificó el fragmento mayor (5524 pb) que corresponde al vector completo menos los 37 pb del gen *Rep* comprendido entre los sitios de corte. El fragmento mayor se usó como vector para insertar cada uno de los ocho casetes sintéticos de doble cadena (figura 14A). Las ligaciones se transformaron en *E. coli* TOP10, y las clonas recombinantes obtenidas para cada uno de los casetes sintéticos, fueron analizadas primero por patrones de digestión utilizando el corte doble *Apal/NgoMIV* (deben producir 3 bandas con los siguientes tamaños: 2960 pb, 1764 pb y 837). En la figura 14C se muestra el patrón de bandas producido por digestión de las clonas que corresponden a los casetes [aa 5,8,10] KKQ, NQK Y KLN; las clonas que presentaron las bandas esperadas son resaltadas con rectángulos rojos, y se enviaron a secuenciar; este mismo procedimiento se aplicó para las clonas que corresponden a los casetes KRN, KKS, KQN y GSK. Las clonas mutantes obtenidas fueron designadas del siguiente modo: M1-KRN (“tipo PHYVV”), M2-KKS (“tipo CdTV”), M3-KKQ (“tipo ACMV”), M4-GSK (“tipo

SiMSinV”), M5-KLN (“tipo TYLCThV”), M6-7-KQN (“tipo OkYMMV/ TGMV”, respectivamente) y M8-NQK (“tipo BCTV”).

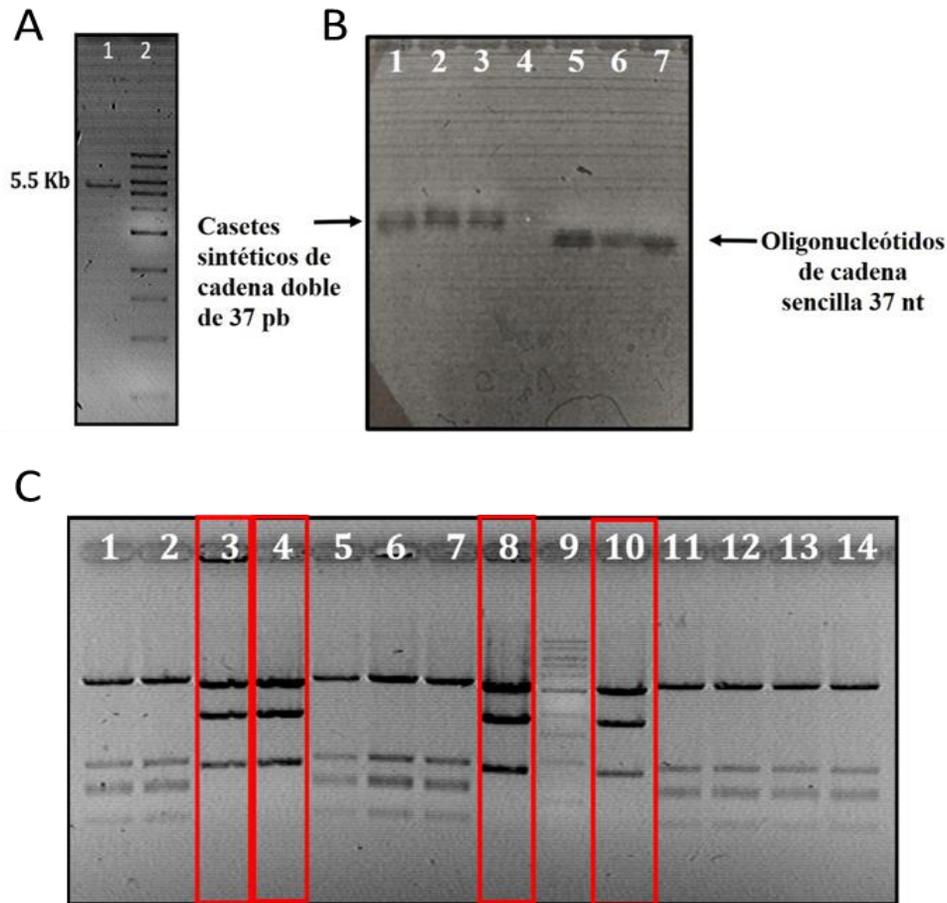


Figura 14. Clonación de casetes sintéticos en componente A de ToMov. (A). gel de agarosa al 1%, pBSK-ToMoVA-NgoMIV tratada con *XmnI*/*NgoMIV* **(B).** gel de agarosa al 3%, línea 1, 2 y 3 corresponden a los casetes sintéticos de doble cadena KKQ, NQK y KLN, línea 5 a la 7 son oligonucleótidos de 37 nucleótidos cadena sencilla. **(C).** Digestiones de las clonas con *ApaI*/*NgoMIV*, carril 1 KKQ-6, carril 2 KKQ-7, carril 3 KKQ-8, carril 4 KKQ-9, carril 5 KKQ-10, carril 6 NQK-6, carril 7 NQK-7, carril 8 NQK-10, carril 9 MPM, carril 10 KLN-6, carril 11 KLN-7, carril 12 KLN-8, carril 13 KLN-9, carril 14 KLN-10. Se resaltan las clonas que mostraron el patrón de digestión esperado, y fueron verificadas por secuenciación.

Generación de variantes de ToMoV-A con mutaciones en los codones 69 y 71 del gen Rep.

La proteína Rep de Pepper huasteco yellow vein virus (PHYVV) y del Chino del tomate virus (CdTV), contienen residuos aa 69 y 71 que son idénticos a los de ToMoV (Q y T, respectivamente). En consecuencia, no fue necesario introducir más cambios en el gen Rep de las variantes “tipo” PHYVV y CdTV que se generaron. Sin embargo, los

restantes 6 mutantes del DNA-A de ToMoV requerían modificar uno o dos de los codones mencionados. Para hacer esto se utilizaron como “plantillas” los DNA-A con las mutaciones correspondientes en los codones 5, 8 y 10 y se procedió a mutar sus codones 69 y 71 utilizando la técnica de PCR mutagénico (ver Materiales y Métodos) siguiendo las recomendaciones del proveedor. En la Tabla Complementaria S2 (Anexos) se indican las secuencias de los oligonucleótidos mutagénicos utilizados en cada caso, y en la Figura S4 (Anexos) se muestra un diagrama simplificado de la metodología usada para el PCR-mutagénico de los codones 69 y 71 del gen *Rep*, utilizando un ejemplo específico. Las mutaciones fueron confirmadas por secuenciación. De esta manera, los componentes derivados del DNA-A de ToMoV, con los genes *Rep* modificadas en los 5 codones diana (5, 8, 10, 69 y 71) fueron obtenidos y designados del siguiente modo: M3 KKQ/VT (DEs tipo *African cassava mosaic virus*, ACMV), M4-GSK/NT (DEs tipo *Sida yellow mosaic Sinaloa virus*, SiMSinV), M5-KLN/KQ (DEs tipo *Tomato yellow leaf curl Thailandia virus*, TYLCThV), M6-KQN/KQ (DEs tipo *Okra yellow mosaic Mexico virus*, OKYMMV), M7-KQN/CQ (DEs tipo *Tomato golden mosaic virus-TGMV*) y M8-NQK/QR (DEs tipo *Beet curly top virus*, BCTV), lo cual se resume en la Tabla 3.

Tabla 3. DNA-A mutantes en los codones que corresponde a los DE.

Mutante	Determinantes de Especificidad	Virus donde se presenta la combinación de DES
M1-PHYVV	KRN/QT	PHYVV
M2-CdTV	KKS/QT	CdTV
M3-ACMV	KKQ/VT	ACMV
M4-SiMSinV	GSK/NT	SiMSinV
M5-TYLCThV	KLN/NT	TYLCThV
M6-OKYMMV	KQN/KQ	OKYMMV
M7-TGMV	KQN/CQ	TGMV
M8-BCTV	NQK/RT	BCTV

Los virus mutantes verificados por secuenciación fueron utilizados para realizar un alineamiento de los primeros 80 residuos aminoácidos del ORF AC1 (*Rep*) utilizando el software Clustal W. Los presuntos determinantes de especificidad de las proteínas *Rep* mutantes se indican en letras en rojo o azul en la Figura 15.

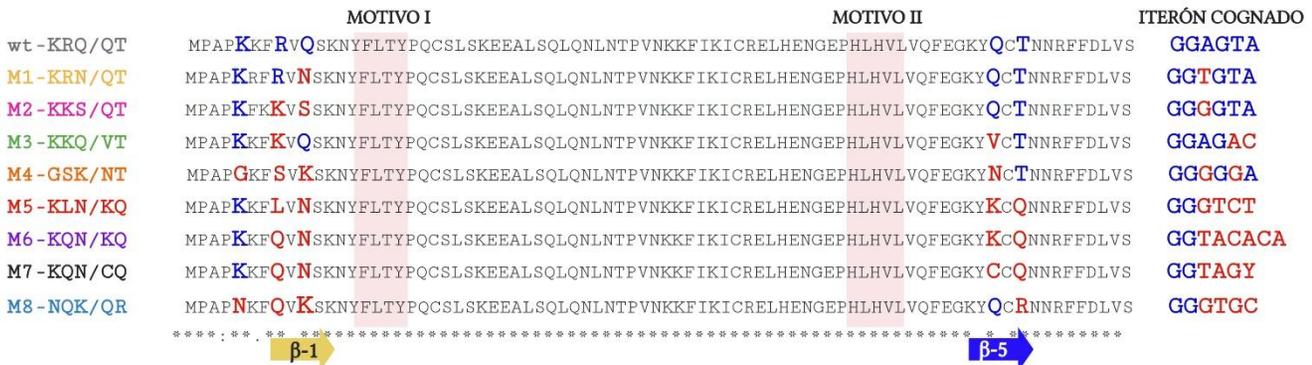


Figura 15. Alineamiento de los primeros 80 aa de las proteínas Rep mutantes y silvestre de ToMoV. A la derecha se encuentra la secuencia del Iterón cognado (predicho). En recuadros magenta se resaltan los motivos conservados RCR I y II. Las flechas representan a los elementos estructurales que forman parte de la mini-lámina beta1,5. La flecha amarilla representa a la hebra-β1, la azul representa a la hebra-β5. El IRD comprende el segmento 5-11 de la proteína.

Ensayos de autoreplicación de los componentes A mutantes de ToMoV en protoplastos de tabaco.

Los componentes A mutantes obtenidos hasta esta fase preliminar fueron examinados en ensayos de autoreplicación usando células en suspensión de *Nicotiana tabaccum* NT1. Las células fueron sometidas a tratamiento con una solución enzimática que degrada parcialmente la pared celular para generar protoplastos (ver Materiales y Métodos). Estos fueron transformados por electroporación con 10 µg DNA de cada uno de los componentes A mutantes, es decir, con DNA-A que codifica diferentes versiones de Rep. El DNA extraído de los protoplastos fue digerido con las enzimas de restricción *DpnI* y *ApaI*. La primera endonucleasa solo reconoce el sitio de corte si éste se encuentra metilado, es decir si el DNA proviene de las bacterias y por lo tanto nos permite discriminar entre el DNA electroporado, del que fue sintetizado “*de novo*” en la célula vegetal. La enzima *ApaI* nos permite linearizar el genoma viral. Posteriormente el DNA total fue fraccionado por electroforesis en gel y se transfirió a una membrana de nylon para realizar un Southern blot usando quimioluminiscencia, como se describe en Materiales y Métodos. La sonda usada fue un fragmento *SacI/ApaI* de 548 pb que corresponde a un segmento del gen *Rep*.

En la figura 16 se muestra el auto-radiograma resultante; los controles positivos fueron los componentes A silvestre (pBSKΔXN-ToMoV-A) y el componente A modificado con un sitio *NgoMIV* (pBSKΔXN-ToMoV-A^{NgoMIV}), estos se muestran en los carriles 1 y 2 respectivamente. Como se puede apreciar, la introducción del sitio *NgoMIV* no alteró el nivel de replicación de ToMoV wt. En el carril 12 se muestra el control negativo, el vector pBSKΔXN vacío. En los carriles del 3 al 10 se colocaron los componentes A

mutantes. Como se observa en la **Figura 16** en ambos controles positivos se detectaron 2 isoformas de DNA, la forma de doble cadena lineal que corresponde a un tamaño de ~2.6 Kb (componente A linealizado) y una banda que migra en un tamaño aparente de 1.4 Kb, que corresponde a la forma de vDNA de cadena sencilla. Los carriles 3, 4, 5, 6, y 8 corresponden a las mutantes tipo M1/PHYVV, M3/ACMV*, M2/CdTV*, M4/SiMSinV, M5/OKYMV respectivamente, y como se observa en la autorradiografía la detección de isoformas de DNA viral sintetizado es mucho menor que los detectados en los controles positivos, lo cual indica un reconocimiento muy débil de los iterones nativos de ToMoV, que resulta, en un nivel de replicación muy bajo. En los carriles 9 y 10, que corresponden a las mutantes tipo M7/TGMV y M8/BCTV respectivamente no se detectó señal indicativa de síntesis de DNA *de novo* de doble o cadena sencilla, en comparación con los controles positivos. Esto sugiere que esas mutantes perdieron prácticamente toda capacidad de unir a los iterones de ToMoV (silvestre).

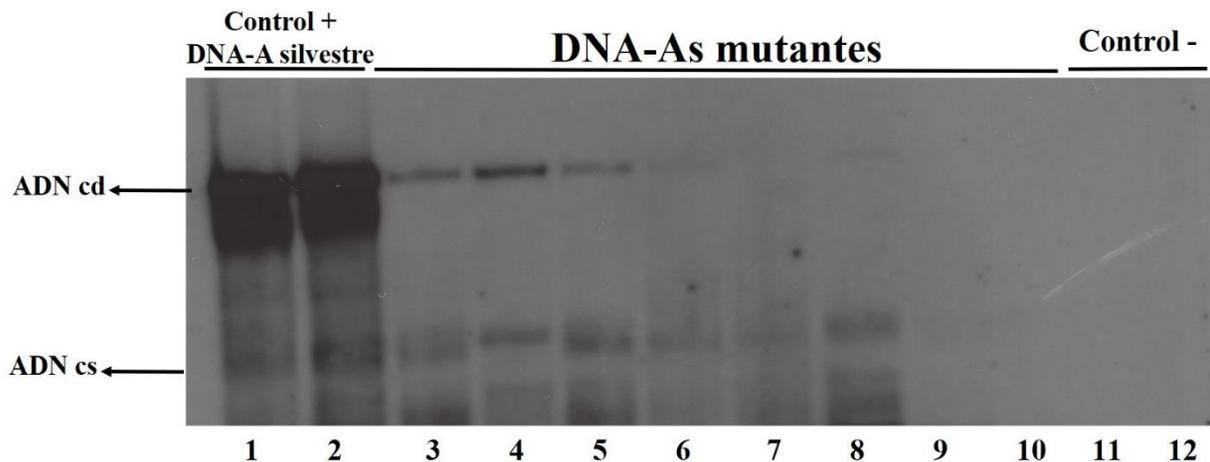


Figura 16. Ensayos de Auto-Replicación de componentes mutantes de ToMoV en protoplastos NT1. Autorradiografía del Southern blot, carriles 1 y 2, controles positivos (ToMoV-A y ToMoV-A^{NgomIV}), carriles 3 al 10 Componentes A mutantes en Rep, Carril 3, variante PHYVV; carril 4, variante ACMV; carril 5, variante CdTV; carril 6, variante SiMSinV; carril 7, variante TYLCThV; Carril 8, variante OkYMMV; Carril 9, variante TGMV; Carril 10, variante BCTV; Carriles 11 y 12 controles negativos. ADNdc, DNA de cadena doble, ADNcs, DNA de cadena sencilla.

Componentes genómicos A doble-mutantes.

El ensayo de Southern blot demostró que las proteínas que presentan una combinación de DES diferentes a los de la proteína Rep nativa, disminuyen drásticamente o pierden por completo su capacidad de reconocer a los iterones silvestres de ToMoV. Sin embargo, dado que nuestro objetivo es definir cuáles son los residuos aminoácidos que determinan su unión preferente a su secuencia cognada, resultaba necesario que los

componentes con los genes Rep mutantes incluyeran, en su origen de replicación, a los iterones que podrían ser reconocidos y unidos por la proteína mutante correspondiente, y determinar si esos dobles mutantes son capaces o no de autorreplicarse. En otras palabras, deben generarse nuevos componentes virales con elementos en *cis* y en *trans* congruentes con el modelo predictivo. Como se mencionó antes, los iterones son secuencias ubicadas en la región intergénica entre el inicio del gen *Rep* y la secuencia donde inicia la replicación por círculo rodante, y funcionan como sitios de unión de alta afinidad de la proteína Rep, reprimiendo la transcripción de su propio gen (Eagle y Hanley-Bowdoin, 1994) y favoreciendo la replicación del vDNA. En consecuencia, en la siguiente fase de nuestro proyecto sustituimos los iterones nativos de ToMoV por los elementos correspondientes a los predichos para cada una de las proteínas Rep mutantes generadas, respetando el arreglo de los cuatro iterones que presenta ToMoV en su origen de replicación.

La región intergénica de ToMoV-A fue por ello utilizada como referencia para diseñar parejas de oligonucleótidos extensos (94 nt) que son complementarios en un segmento de ~40 nt que, mediante extensión bidireccional con una DNA polimerasa que usa de molde la cadena parcialmente complementaria, se fusionen formando un segmento sintético de doble cadena de aproximadamente 135 pb (ver Materiales y métodos). Este segmento de DNA contiene, de acuerdo con el diseño original, un extremo que contiene un sitio de corte para la enzima *PmlI*, y un extremo opuesto que contiene un sitio *NgoMIV*, por lo cual, una vez digerido con dichas enzimas el segmento de ~135 pb puede ser entonces ligado a los vectores ToMoV-A mutantes, reemplazando al segmento (*PmlI/NgoMIV*) que contiene los iterones nativos de ToMoV, con una secuencia de igual tamaño que contendrá iterones diferentes, que corresponderán a los predichos para cada proteína Rep específica. En la Figura 17 se muestra la secuencia de un par de los macroligonucleótidos diseñados para introducir iterones de tipo OKYMMV (GGTACACA) para reemplazar los iterones de ToMoV (GGAGTA), donde se indican también los sitios *PmlI* y *NgoMIV* indispensables para hacer el reemplazo de la región nativa de ToMoV, de tal modo que la mutante Rep de “tipo OKYMMV” queda dotada de un origen de replicación del mismo tipo, dando así lugar al ToMoV doble-mutante con esa denominación (ver más adelante).

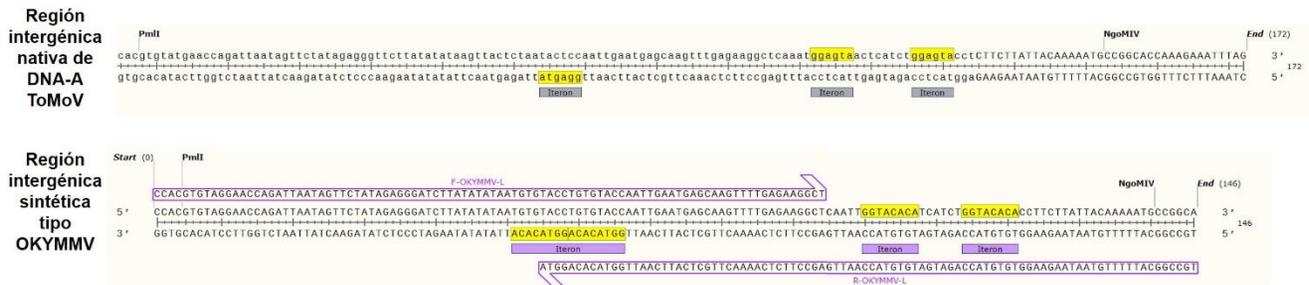


Figura 17. Diseño de oligonucleótidos extensos. Se diseñaron 8 parejas con iterones diferentes a los nativos en ToMoV. Cada secuencia se enlista en la Tabla S1. En este caso, los iterones de ToMoV (GGAGTA) son sustituidos por los del tipo OKYMMV (GGGTGC)

La **Figura 18** es una representación esquemática general, del reemplazo de la secuencia que corresponde a los iterones silvestres (**GGAGTA**) de la mutante ToMoV M8-BCTV, por un segmento que tiene los iterones cognados (**GGGTGC**) que correspondan a los DES de la proteína Rep tipo BCTV. Este mismo proceso se realizó para los 8 componentes A mutantes. Cada componente A mutado fue verificado por secuenciación. Adicionalmente, cada secuencia se analizó con el software SnapGene® para verificar que el origen de replicación correspondiese a las mutaciones realizadas previamente en el gen *Rep*.

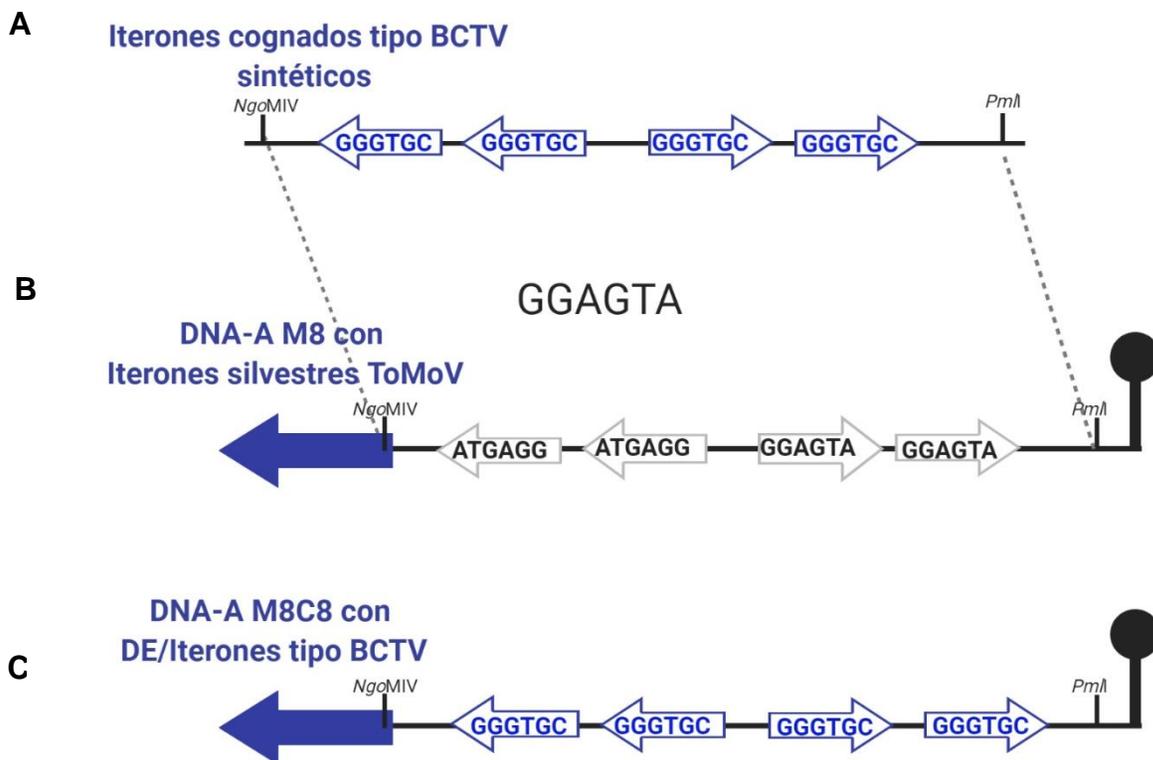


Figura 18. Construcción de los DNA-A dobles mutantes. (A). Representación esquemática del reemplazo del segmento nativo *Pml/NgoMIV* (145 pb) río arriba del gen *Rep* por los segmentos sintéticos. (B) OriV wt de ToMoV. (C). Rearreglo de iterones en el oriV mutante M8C8. Para efectos de claridad, los iterones invertidos (CCCACG y CCTCAT) se representaron con la misma secuencia que los iterones en la cadena de polaridad positiva.

los aminoácidos KKQ/VT en Rep e iterones de ACMV (virus de mosaico africano de la cassava) y el último número corresponde al número de clona. En la tabla 4 se describe cada una de las dobles mutantes

Tabla 4. Componentes A dobles mutantes (DE/ iterones cognados)

Mutante	Determinantes de Especificidad	Iterón	Virus donde se presenta la combinación
M1P1	KRN/QT	GGTGTA	PHYVV
M2C2	KKS/QT	GGGGTA	CdTV
M3A3	KKQ/VT	GGAGAC	ACMV
M4S4	GSK/NT	GGGGGA	SinMSiV
M5T5	KLN/KQ	GGGTCT	TYLChV
M6O6	KQN/KQ	GGTACACA	OkYMMV
M7T7	KQN/CQ	GGTAGy	TGMV
M8C8	NQK/RT	GGGTGC	BCTV

Análisis de los componentes A doble-mutantes

Una vez generados los ToMoV-A doble-mutantes, la siguiente etapa del estudio consistió en evaluar la capacidad replicativa de cada uno de los componentes genómicos A mutantes de ToMoV, los cuales codifican una proteína Rep con los presuntos determinantes de especificidad afines (cognados), y con un origen de replicación que contiene iterones distintos a los nativos, pero similares a los predichos para la proteína mutante correspondiente. En un ensayo anterior de este estudio usamos con éxito protoplastos de *N. tabacum* para examinar la replicación de ToMoV, tanto el virus silvestre como el mutante con el sitio NgoMIV al inicio del gen *Rep* (ver Fig 16). Intentamos repetir este tipo de ensayo utilizando los dobles mutantes de ToMoV, pero esta vez no logramos obtener replicación transitoria, no solo con los nuevos mutantes, sino con los mismos controles usados en el ensayo anterior (ToMoV-A wt y ToMoV-A^{NgoMIV}). Después de repetidos intentos y cambio de condiciones, llegamos a la conclusión que la línea celular NT1 original se había alterado de algún modo (ej. fisiológica o epigenéticamente), pues incluso se observaron aglomerados celulares anómalos bajo ciertas condiciones.

Esta situación nos obligó a buscar metodologías alternativas para examinar la auto-replicación de las mutantes y poner a prueba la hipótesis del trabajo. Por ello, optamos por realizar ensayos de auto-replicación transitoria en disco de hoja de *N. benthamiana* (Kim *et al.*, 2007). El sistema consiste en agroinfiltrar hojas con una suspensión de *Agrobacterium tumefaciens* portadoras de clonas infectivas de los virus, que son hemidímeros de los componentes genómicos A tanto el silvestre como los mutantes, clonados en un vector binario. Los hemidímeros se caracterizan por incluir un

componente viral (A o B) completo flanqueado por dos orígenes de replicación idénticos.

Por la complejidad del diseño y las dificultades metodológicas asociadas, decidimos trabajar tan solo con 5 de los 8 componentes A mutantes previamente generados, por lo que seleccionamos a los siguientes cinco mutantes del componente A de ToMoV: del tipo PHYVV (**M1P1**), ACMV (**M3A3**), OKYMMV (**M6O6**), TGMV (**M7T7**) y BCTV (**M8B8**). Estas combinaciones son representativas, e incluyen variantes con DEs diferentes e iterones con secuencias marcadamente distintas de las de ToMoV. A continuación, describimos en detalle la construcción y obtención de las clonas infectivas del componente A de las mutantes mencionadas.

Clonas infectivas de los componentes A mutantes y silvestre.

Para la construcción de las clonas infectiva del DNA-A (wt y mutantes), se amplificó un segmento viral de 1388pb de cada monómero mutante usando los iniciadores F1THK/M13-Rev; este fragmento, que incluye la región intergénica más 837 pb del gen *Rep*, se clonó en el vector pGEM-T Easy y se secuenció. Una vez verificada por secuenciación, la clona fue digerida con las enzimas *EcoRI* y *HindIII*, para liberar el fragmento de aproximadamente 1.3 Kb (Figura 18B), el cual se subclonó en el vector binario pCambia 1300; las clonas fueron analizadas por patrón de digestión usando *Apal/Pml* y posteriormente se enviaron a secuenciar (más detalles en anexos). La construcción pCa+1.3Kb se usó como vector para clonar el monómero (DNA-A mutante o silvestre según corresponda) en un sitio *Apal* y así se obtuvieron, en cada caso, los monómeros de DNA-A silvestre y mutantes, flanqueado por dos orígenes de replicación similares (Figura 20A).

De esta forma se lograron construir seis clonas infectivas que corresponde a ToMoV-A^{NgoMIV}, y las mutantes M1P1 (“PHYVV”), M3A3 (“ACMV”), M6O6 “OKYMMV”, M7T7 “TGMV”, y mutante M8C8 “BCTV”. Por último, estas construcciones de transformaron por electroporación en *A. tumefaciens* GV3101. Las células transformadas se identificaron por su resistencia a los antibióticos Km, Rif y Gm, y posteriormente por PCR en colonia usando oligonucleótidos degenerados específicos para begomovirus DGR-For/YMAC-Rev (replicón de aproximadamente 1400 pb), Figura 20B. Luego de verificar por secuenciación, se seleccionaron las clonas que se habrían de utilizar para realizar los ensayos de replicación y/o infección. En la tabla 5 se enlistan las clonas de *A. tumefaciens* GV3101 obtenidas en este trabajo.

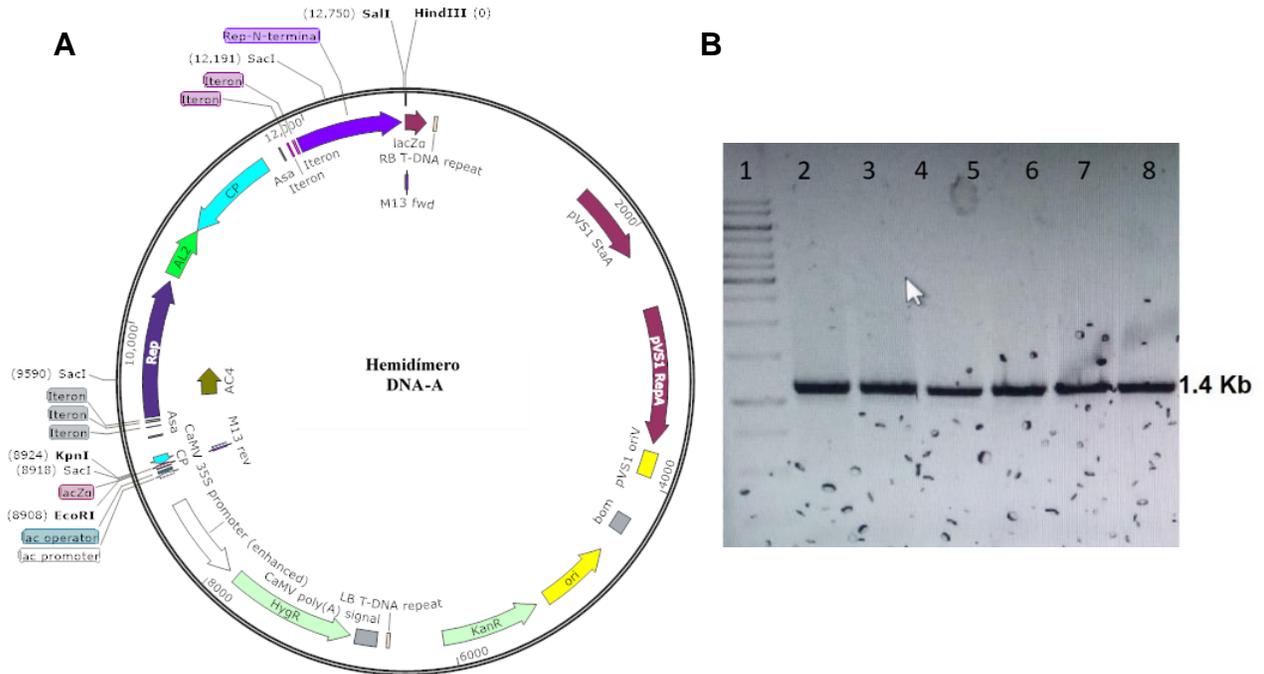


Figura 20. Mapa físico de las clonas infectivas de DNA-A de ToMoV-A mutante o silvestre. (A). Mapa físico de los hemidímeros del componente A, en el vector pCAMBIA1300. **(B).** PCR en colonia de *A. tumefaciens* GV3101 transformadas con las clonas infectivas, carril 1 MPM, carril 2. *A.tumefaciens*-MIPI, carril 3 *A.tumefaciens*-M3A3, Carril 4 *A.tumefaciens*-M5T5, carril 5 *A.tumefaciens*-M6O6, carril 6 *A.tumefaciens*-M7T7, carril 7 *A.tumefaciens*-M8C8, carril 8 control positivo *A.tumefaciens*-ToMoV-A wt.

Tabla 5. Clonas infectivas de componentes A mutantes y silvestre

Nombre	Construcción	Virus nativo representativo
<i>A.tumefaciens</i> -ToMoV-A wt.	pCambia+hemidímero de DNA-A silvestre (13 Kb)	ToMoV
<i>A.tumefaciens</i> -ToMoV- <i>A</i> _{NgoMIV}	pCambia+hemidímero de DNA-A con un sitio NgoMIV (13 Kb)	ToMoV
<i>A.tumefaciens</i> -M1P1	pCambia+hemidímero de DNA-A mutante tipo PHYVV (13 Kb)	PHYVV
<i>A.tumefaciens</i> -M3A3	pCambia+hemidímero de DNA-A mutante tipo ACMV (13 Kb)	ACMV

<i>A.tumefaciens</i> -M6O6	pCambia+hemidímero de DNA-A mutante tipo OKYMMV (13 Kb)	OKYMMV
<i>A.tumefaciens</i> -M7T7	pCambia+hemidímero de DNA-A mutante tipo TGMV (13 Kb)	TGMV
<i>A.tumefaciens</i> -M8C8	pCambia+hemidímero de DNA-A mutante tipo BCTV (13 Kb)	BCTV

Ensayos de autoreplicación en discos de hoja de *N. benthamiana*

Las clonas infectivas de los componentes genómicos A silvestre y doble mutantes se examinaron (por triplicado) en ensayos de auto-replicación en disco de hoja de *N. benthamiana* (Kim, 2007), con la finalidad de probar si las proteínas Rep mutantes son capaces de reconocer y unir sus presuntos iterones cognados. El sistema de transferencia de DNA a planta que se utilizó fue el sistema de Agroinfiltración (Figura 19) como se describe en detalle en Materiales y métodos.

Se utilizaron las cepas de *Agrobacterium* portadoras de las clonas infectivas de los componentes A mutantes (MIPI, M3AC, M6O6, M7T7 Y M8C8, y el virus silvestre ToMoV-A^{Ng^oMIV} (Figura 21). El buffer de agroinfiltración se aplicó en la superficie abaxial de hojas de plantas de 4 semanas de edad; posteriormente se dejaron incubar por 4-5 días en una cámara ambiental a 25 °C con ciclo de 12 horas de luz/obscuridad, período en el que se espera ocurra la auto-escisión del monómero viral, el cual se replicará usando el mecanismo de RCR, como se muestra en la Figura 18A. Las hojas agroinfiltradas se usaron para extraer el DNA total, se transfirieron 50 µg de DNA tratado con las enzimas *DpnI*/*Apal* y una vez fraccionado por electroforesis en gel se transfirió el DNA a una membrana de nylon, la cual se trató usando un protocolo bien establecido de Southern blot, usando como sonda un fragmento del gen *Rep*, utilizado en los ensayos anteriores (figura 21B)

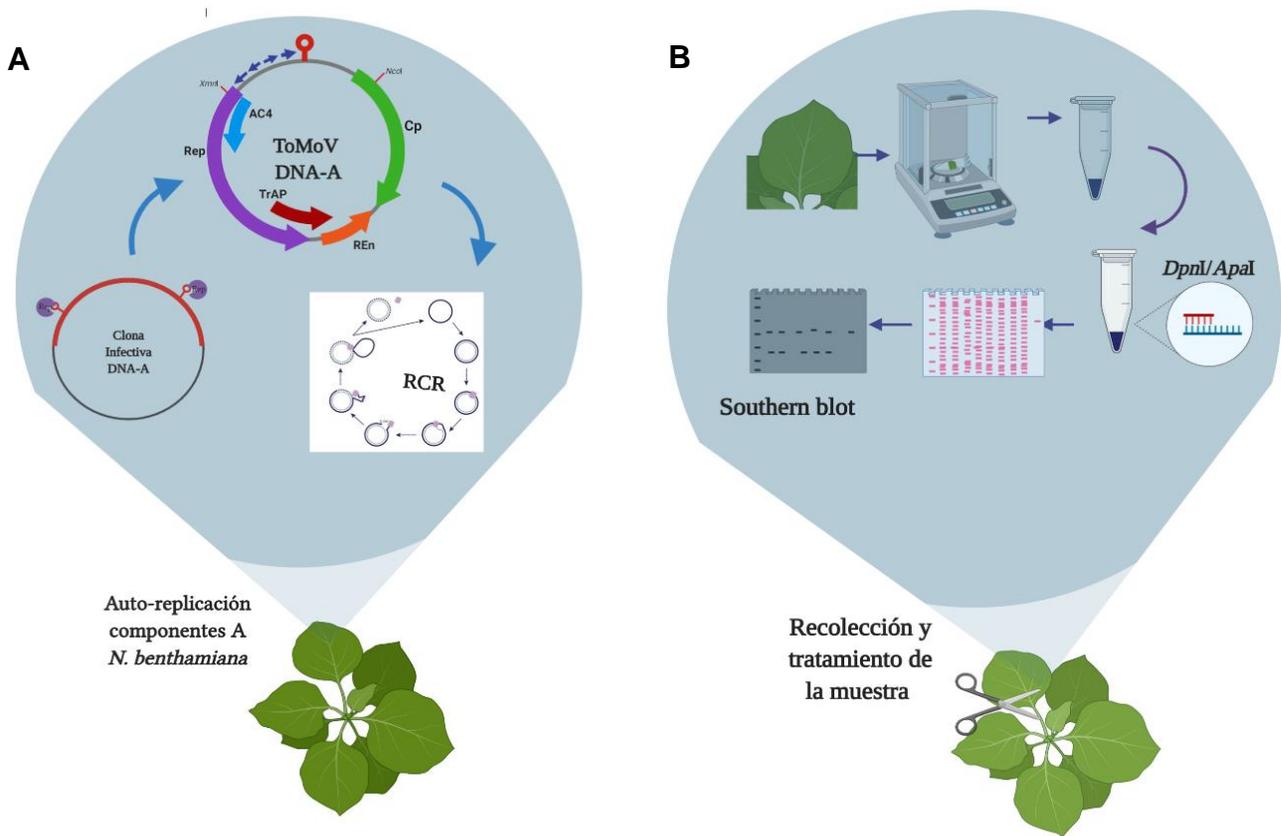


Figura 21. Diagrama ilustrativo de ensayos de auto-replicación en discos de hoja. (A). Ilustración del proceso de Agroinfiltración al usar clonas infectivas de componentes geminivirales. (B). Recolección y pasos de tratamiento de la muestra para la detección y análisis del DNA viral. (Biorender.com)

En la **Figura 22** se muestra la auto-radiografía del Southern blot obtenido después de agroinocular una o dos hojas de plantas individuales de *Nicotiana benthamiana* con clonas infectivas del componente genómico A de ToMoV (control positivo) y cinco dobles mutantes (en Rep y los iterones) del DNA-A derivado de ToMoV. El carril 1 corresponde al control positivo, ToMoV^{NgomIV}, el carril 2 no contiene DNA, los carriles del 3 al 7 corresponden al DNA de las 5 dobles mutantes del DNA-A, indicados en el pie de figura, y los carriles 8 y 9 corresponden a DNA de hoja sin infiltrar, y DNA agroinfiltrada sólo con el buffer de infiltración ("mock"), respectivamente. En el carril 10 se cargó DNA plasmídico de ToMoV (50 ng) linealizado con *ApaI*, como referencia. Se observa claramente la presencia de DNA *de novo* de doble cadena (dsDNA) que comigra con el DNA plasmídico, y otras formas de DNA de cadena sencilla (ssDNA) en el control positivo silvestre (ToMoV-^{NgomIV}), y en las dobles mutantes, debido al reconocimiento de los iterones por Rep, tal como ocurre en la naturaleza.

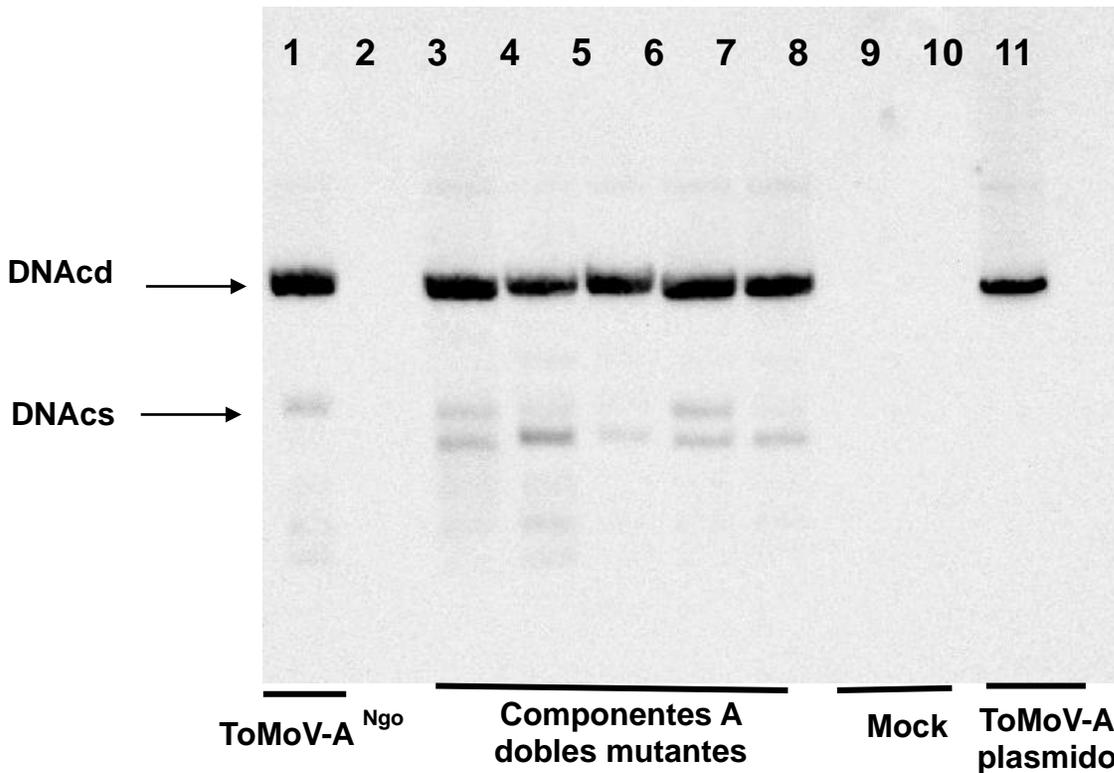


Figura 22. Ensayo de auto-replicación en discos de hoja. Southern blot con DNA obtenido de hojas agroinfiltradas con los componentes A de ToMoV silvestre y mutantes. (C-1) ToMoV-A^{NgoMIV} (carril-2) vacío; (carril 3), DNA de hoja agroinfiltrada con la clona mutante **MIPI-PHYVV**; carril 4, DNA de hoja infiltrada con la clona **M3A3-“ACMV”**; carril 5, DNA de hoja agroinfiltrada con la clona **M6O6**, OkYMMV; carril 6, hoja infiltrada **M7T7**, TGMV; carril 7 DNA de hoja agroinfiltrada con la clona **M8B8**, “BCTV”; carril 8, DNA de hoja sin infiltrar; carril 9, DNA de hoja infiltrada con el buffer de agroinoculación (“mock”); Carril 10, DNA-A plasmídico (linearizado con Apal), 2.6 Kpb. La membrana fue revelada usando quimioluminiscencia, la imagen fue capturada en el equipo ChemiDoc MP Imaging System (BIO-RAD) y en el procesamiento de imagen se usó Image **Lab Software (BIO-RAD)**.

Construcción de componentes B de ToMoV con iterones diferentes.

Se realizó la construcción de componentes B mutantes previamente clonados en el vector pBSKΔXE, por medio del intercambio de un fragmento nativo *EcoRI/PmlI* de 345 pb, ubicado en la región intergénica del componente genómico B de ToMoV silvestre. Esta región contiene los iterones de ToMoV y sus mutantes derivados por el fragmento sintético que se utilizó para el intercambio de iterones del componente genómico A, como se detalla en Materiales y métodos. Dado que el segmento *EcoRI/PmlI* de los mutantes A es menor (~158pb), los DNA-B mutantes obtenidos son cerca de 180pb más pequeños que el DNA-B silvestre. El segmento eliminado, corriente arriba del gen *BC1*, no parece participar en el proceso de replicación, y no contiene, probablemente, elementos *cis*-reguladores de la transcripción (datos no mostrados). Los componentes

B mutantes: en su configuración infecciosa (hemidiméricos) son **DNA-B1** (iterones tipo PHYVV), **DNA-B3** (iterones tipo ACMV), **DNA-B6** (iterones tipo OKYMMV), **DNA-B7** (iterones tipo TGMV) y **DNA-B8** (iterones tipo BCTV) (**Figura 23**), cada uno de ellos verificados por patrón de digestión, PCR y por secuenciación.

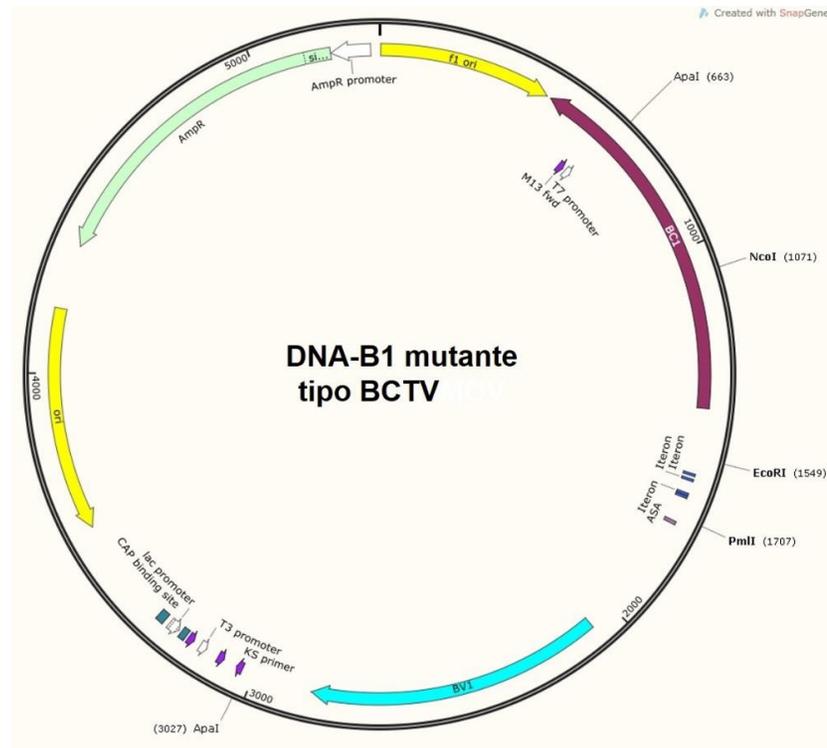


Figura 23. Mapa físico de DNA-B1 mutante **tipo BCTV**. El componente B mutante tipo BCTV fue clonado en el vector pBSKΔXE.

Construcción de clonas infecciosas de componentes genómicos B con iterones diferentes a ToMoV

Para los ensayos de infección realizados en plantas, es necesario que los componentes genómicos B se utilicen en su forma infecciosa, es decir, en forma de dímeros o hemidímeros. Los hemidímeros se construyeron en el vector binario pCAMBIA1300. Un fragmento *EcoRI*/*SacI* de 1751 pb (correspondiente a la región intergénica y 774 pb del gen BV1) de cada monómero mutante o silvestre, se clonó en sitios compatibles del vector binario pCAMBIA, dando origen a una construcción pC+Hemi-1.7, con un tamaño aproximado de 10,700 pb; las clonas se analizaron por patrón de digestión y posteriormente se enviaron a secuenciar. El siguiente paso fue clonar en cada construcción el monómero (2.6 Kb) del DNA-B completo

correspondiente, en un sitio *ApaI*; las clonas resultantes fueron confirmadas por sus patrones de digestión para verificar si la construcción estaba en la orientación correcta, como se detalla en los anexos. Las clonas infectivas (Figura 24) se transformaron en *A. tumefaciens* GV3101. Las células de *Agrobacterium* fueron seleccionadas por la resistencia a los antibióticos Gm, Km y Rf, y además se les realizó una PCR usando oligonucleótidos “universales” para el DNA-B de begomovirus, que se describen en la Tabla 6.

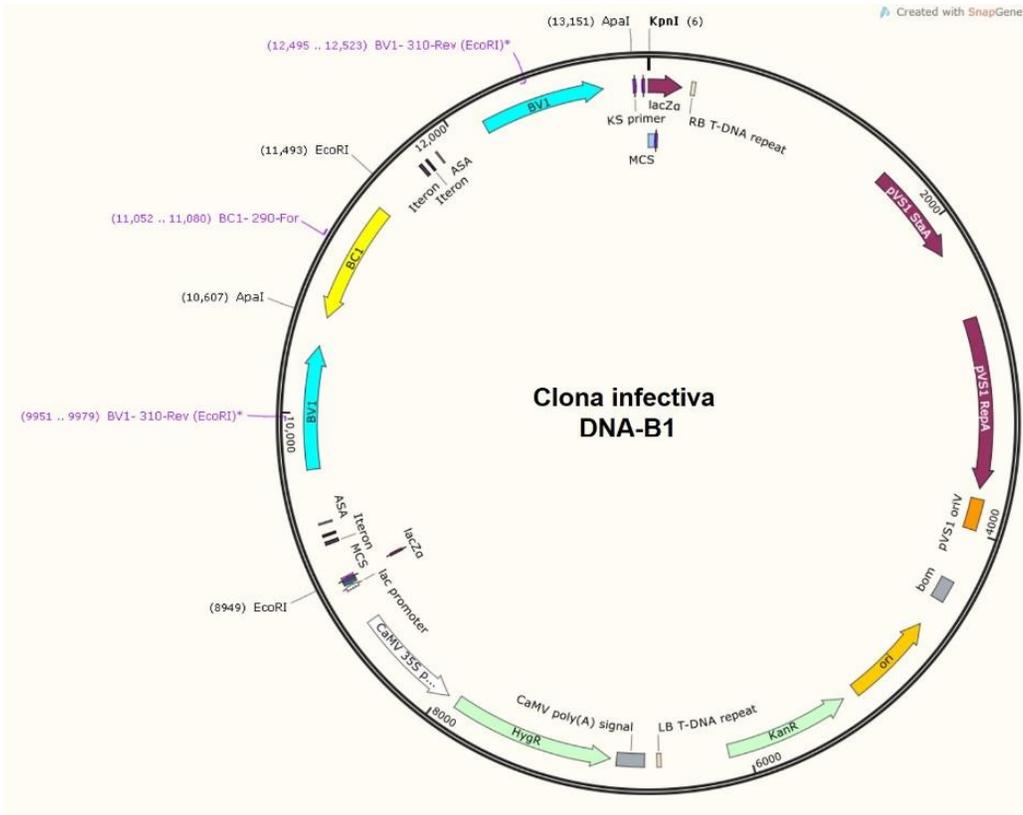


Figura 24. Mapa físico de clona infectiva B1-8C8. Hemidímero de componente B mutante tipo BCTV clonado en un vector binario (pCambia)

Tabla 6 Resumen de las clonas infectivas del DNA-B construidas en este trabajo.

Clona Infectiva	Construcción	Virus silvestre que contiene los iterones
A. tumefaciens GV3101 CI-B wt	pCambia+hemidímero de DNA-B silvestre (13 Kb)	ToMoV
A. tumefaciens GV3101 CI-B1P1	pCambia+hemidímero de DNA-B mutante tipo PHYVV (13 Kb)	PHYVV
A. tumefaciens GV3101 CI-B3A3	pCambia+hemidímero de DNA-B mutante tipo ACMV (13 Kb)	ACMV
A. tumefaciens GV3101 CI-B6O6	pCambia+hemidímero de DNA-B mutante tipo OKYMMV (13 Kb)	OKYMMV
A. tumefaciens GV3101 CI-B7T7	pCambia+hemidímero de DNA-B mutante tipo TGMV (13 Kb)	TGMV
A. tumefaciens GV3101 CI-8C8	pCambia+hemidímero de DNA-B mutante tipo BCTV (13 Kb)	BCTV

Ensayos de Infección en plantas de *N. benthamiana*

Para la infección sistémica de un begomovirus bipartita en plantas, es necesaria la presencia de ambos componentes virales, es decir, el DNA-A y el DNA-B, por lo que se construyeron clonas infectivas del DNA-A y el DNA-B, como se describió previamente. Se realizaron los ensayos de inoculación por triplicado; en cada caso el número de muestras fue de 9 plantas, y para los controles negativos y mock en número de muestras fue de 7 plantas.

Antes de realizar los ensayos de inoculación de interés se realizaron ensayos con controles para verificar si los componentes A doble-mutantes son capaces de reconocer y unir a los iterones cognados presentes en el componente B silvestre de ToMoV. Se usaron 5 plantas por combinación y se realizó el análisis por triplicado. Se recolectaron muestras de las plantas a los 14 y 21 dpi, las cuales, con la excepción de la combinación silvestre, no presentaron síntomas a excepción de los controles positivos que presentan una sintomatología típica de ToMoV en *Nicotiana benthamiana*, como amarillamiento y moteado ligeramente más clorótico en hojas, además de crecimiento retrasado o enanismo (datos no mostrados pero ver figuras más adelante). Se realizó la extracción de DNA total de las plantas agroinoculadas y, por último, se realizó un análisis por PCR para la detección del componente B silvestre usando los iniciadores BCI290-For/BV1310-Rev. Como se muestra en la **Figura 25**, solo se pudo detectar la presencia del componente B silvestre en los controles positivos, lo que indica que **ninguno de los componentes A mutantes pudo reconocer el origen de replicación del componente B nativo de ToMoV-B**, por lo que no fueron capaces de trans-replicarlo y producir una infección sistémica.

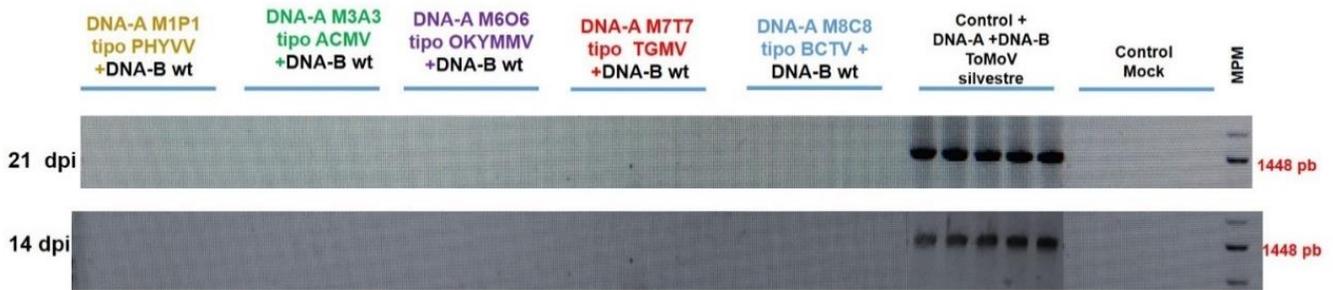


Figura 25. Ensayo de inoculación de componentes A mutantes más componente B silvestre de ToMoV. Productos de PCR solo se observaron en las plantas inoculadas con el DNA-A y el DNA-B silvestres de ToMoV

En los ensayos de inoculación de plantas con DNA-As y DNA-Bs mutantes, se probaron las seis clonas infectivas generadas que corresponden a ToMoV-A silvestre, la mutante **M1PI** (ToMoV-A tipo PHYVV), mutante **M3A3** (“ACMV”), mutante **M6O6** (“OkYMMV”), mutante **M7T7** (“TGMV”), y mutante **M8C8** tipo BCTV, así como cinco clonas infectivas de los componentes B mutantes, correspondientes a cada uno de los DNA-A mutantes antes mencionados. En la **Tabla 7** se ilustra la combinación de cada componente viral como clonas infectivas probadas en las plantas.

Tabla.7 Resumen de combinación *A. tumefaciens* GV3101 utilizadas en los ensayos de infección con ToMoV silvestre y sus derivados mutantes.

Combinación	Tipo de virus de DEs/Iterones	Numero de muestra por ensayo
DNA-A +DNA-B wt	Control positivo	9
MIPI+ B1P1	PHYVV	9
M3A3+B3A3	ACMV	9
M6O6+B6O6	OKYMMV	9
M7T7+B7T7	TGMV	9
M8C8+B8C8	BCTV	9
Mock	Solución de Agroinfiltración	7
Control negativo	No Agroinfiltración	7

En la **Figura 26** se presentan fotografías de las plantas infectadas con los componentes silvestres y los componentes mutantes a los 14 y 21 dpi (días post-inoculación, dpi). Las plantas inoculadas con las construcciones de ToMoV silvestres presentaron síntomas visibles a partir del día 13-14, como el amarillamiento de la hoja con cierto grado de moteado ligeramente más clorótico, y retraso del crecimiento (**Figura 26**). En las plantas inoculadas con las construcciones mutantes se presentaron síntomas similares en las hojas nuevas y el retraso del crecimiento a partir del día 13-14, que fue más evidente a los 21 dpi. El retraso en el crecimiento varió entre las mutantes, con las plantas inoculadas con **M3A3**, **M6O6** y **M8B8** mostrando el mayor retraso en crecimiento respecto a los controles positivos. En la tabla 8 se resumen los resultados de las plantas infectadas. Esos resultados sugieren que las proteínas Rep mutantes fueron capaces de reconocer y unir a los iterones de los componentes A y B modificados, y producir, de ese modo, una infección sistémica al movilizar ambos componentes genómicos a los tejidos nuevos de las plantas.

Tabla 8. Reporte de sintomatología de plantas infectadas

COMBINACION DE CLONAS INFECTIVAS.	SINTOMATOLOGIA	PLANTAS INFECTADAS N=9 por triplicado
Control positivo ToMoV DNA-A/DNA-B silvestre [K5, R8, Q10] [Q69, T71] Iterón cognado: GGAGTA	Amarillamiento de las hojas (con ligero moteado) a partir de los 14 dpi. Crecimiento retrasado a los 21 dpi.	27/27
Combinación: DNA-A MIPI/DNA-B tipo PHYVV DES: [K5, R8, N10] [Q69, T71] Iterón predicho: GGTGTA	Leve amarillamiento, y moteado tenue de hojas, a partir de los 14 dpi, y con retraso en el crecimiento a 21 dpi.	25/27
Combinación: DNA-A M3A3/DNA-B tipo ACMV DES: [K5, K8, Q10] [V69, T71] Iterón predicho: GGAGAC	Amarillamiento en forma de moteado tenue de hojas a los 14 dpi. Retraso en crecimiento a 21 dpi.	27/27
Combinación: DNA-A M6O6/DNA-B tipo OKYMMV DES: [K5, Q8, N10] [V69, T71] Iterón predicho: GGTACAC	Amarillamiento y moteado tenue de hojas a los 14 dpi; mayor retraso en crecimiento a los 21 dpi.	24/27
Combinación: DNA-A M7T7/DNA-B tipo TGMV DES: [K5, Q8, N10] [C69, Q71] Iterón predicho: GGTAGY	Leve amarillamiento y moteado de hojas a los 14 dpi, retraso de crecimiento a 21 dpi.	26/27

Combinación: DNA-A M8C8/DNA-B tipo BCTV DES: [N5, Q8, K10] [Q69, R71] Iterón predicho : GGGTGC	Amarillamiento de hojas, se presenta a los 13-14 dpi y con claro retraso en el crecimiento a 21 dpi	24/27
Mock (Solución de Agroinfiltrado)	Sin síntomas	

B- 21 dpi

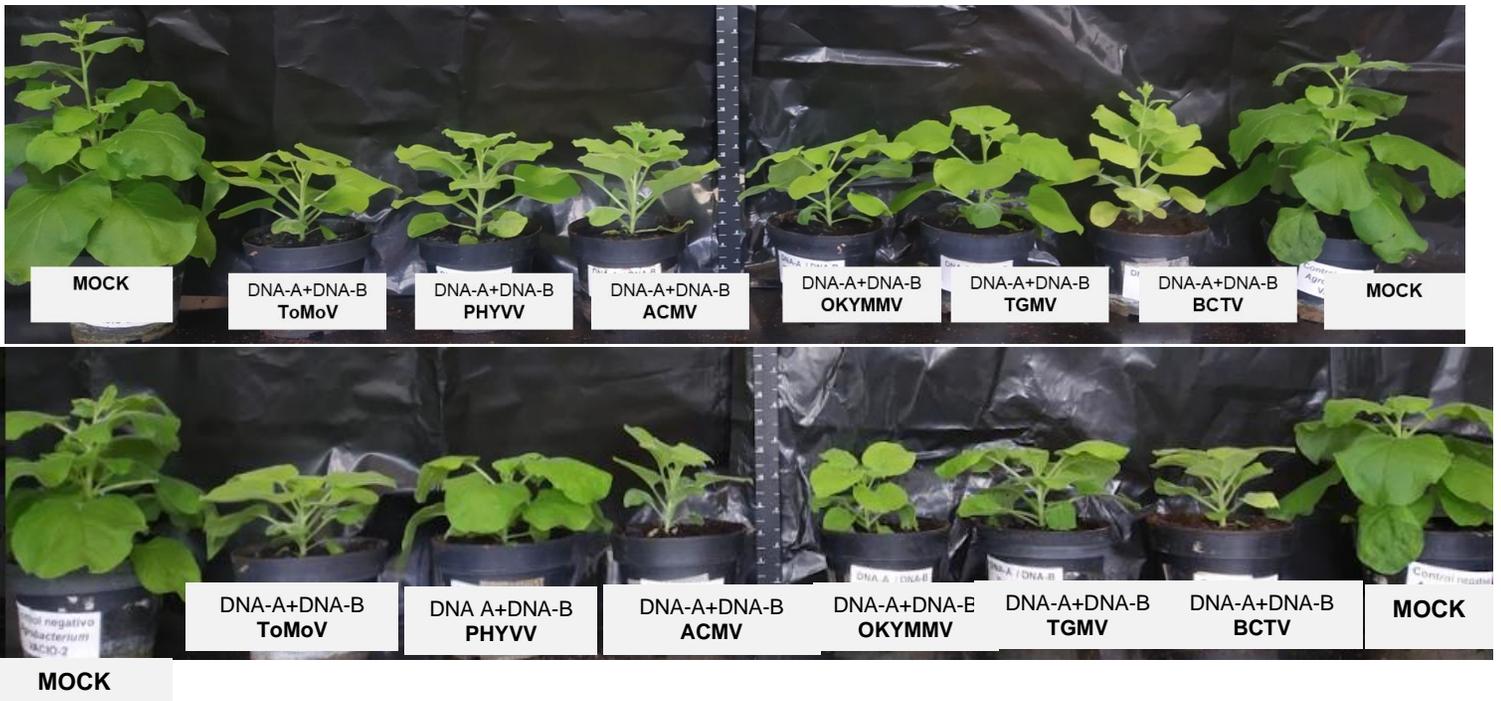


Figura 26. Ensayos de infección de *N. benthamiana*. Fotografías a los 14 y 21 días post-infección. ToMoV y todas sus mutantes mostraron un crecimiento reducido respecto a los controles.

El amarillamiento y el crecimiento menos robusto de las hojas como síntomas de la infección por ToMoV y sus derivados por mutagénesis, se puede apreciar mejor en la **Figura 27**, en la que las fotografías abarcan las hojas más jóvenes de las plantas infectadas y las de una planta control.

VISTA AEREA



DNA-A+DNA-B
ToMoV



DNA-A+DNA-B
PHYVV



DNA-A+DNA-B
ACMV



DNA-A+DNA-B
OKYMMV



DNA-A+DNA-B
TGMV



DNA-A+DNA-B
BCTV



MOCK

Figura 27. Ensayos de infección de *N. benthamiana*. Fotografías de las hojas apicales a los 21 días post-infección. Nótese el amarillamiento de las hojas y estructura más delicada de las plantas inoculadas con los pares infectivos silvestre (ToMoV) y mutantes, respecto al control negativo (mock).

Ensayo de pseudorecombinación 21 dpi

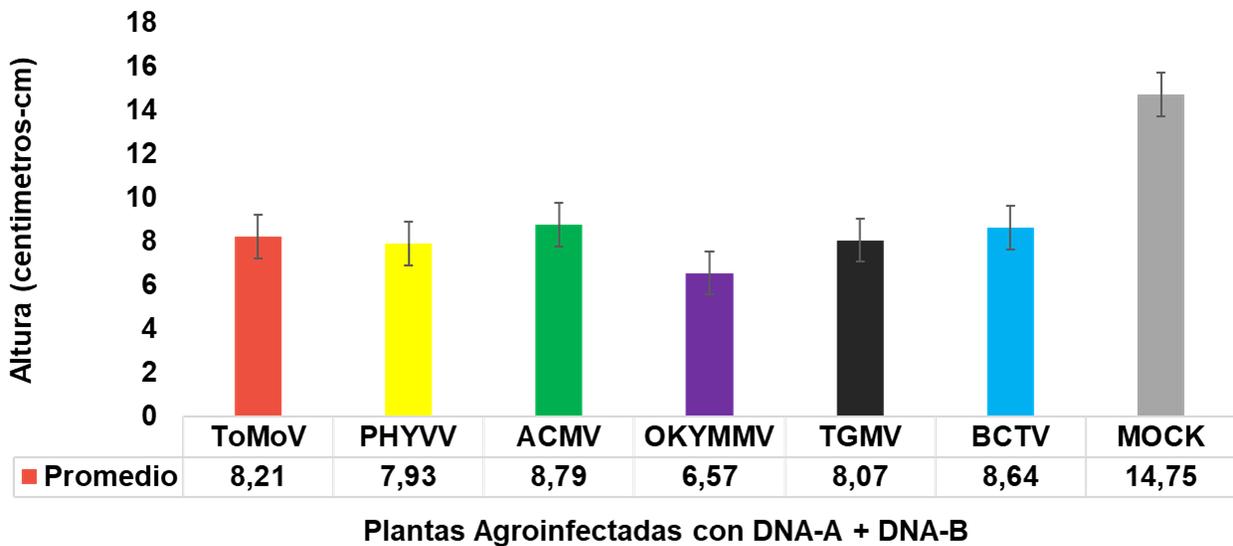


Figura 28-. Gráfica de la altura de plantas en el ensayo de pseudo-recombinación a los 21 pdi. Se agroinfiltraron plantas de *N. benthamiana* con 4 hojas verdaderas, con una combinación de *A. tumefaciens* que portan clonas infectivas de DNA-A y DNA-B de ToMoV y mutantes tipo PHYVVV, tipo ACMV, tipo OKYMMV, tipo TGMV, y tipo Beet curly top virus (BCTV). El grupo que corresponde al control negativo (“mock”) se infiltró con solución de agroinfiltrado mas *A. tumefaciens* GV3101. Prueba estadística para un solo grupo, n=9. Las barras representan el error estándar. El asterisco representa diferencia significativa al comparar los grupos de plantas infectadas con el grupo “mock”, en una prueba de Tukey ($\alpha=0.05$)

Análisis de secuencias de los productos de PCR de plantas infectadas.

A los 14 y 21 dpi, se extrajo DNA total de las hojas más jóvenes de las plantas experimentales, y se amplificó el DNA viral utilizando, en reacciones separadas, iniciadores específicos para el DNA-A y el DNA-B. En todas las plantas sintomáticas se pudieron amplificar los segmentos relevantes de ambos componentes genómicos (Ver figura S1 en Anexo 3). En el caso de los DNA-B mutantes el producto de PCR fue significativamente menor que el DNA-B silvestre, como se explicó previamente (Figura 29). Los productos de PCR de 3 plantas infectadas de cada grupo se clonaron en pGEM T-Easy y se secuenciaron. Los resultados se analizaron por medio de alineamientos para establecer con seguridad que se recuperaran los virus que corresponden a cada ensayo de inoculación. En la **Figura 30** se muestra el alineamiento de secuencias obtenidas de los componentes A silvestre y mutantes. En cada muestra analizada se pudo verificar, primero, que las secuencias corresponden a los componentes virales con los cuales se realizó la inoculación de cada planta; además, que los componentes genómicos presentan las modificaciones realizadas en esta investigación, como son los codones de los DEs que correspondan, y que en su región intergénica se encuentren los iterones cognados (predichos) mismos que se resaltan en el alineamiento con rectángulos de diferentes colores.

Detección de DNA-B con BC1 290-For/BV1-310 Rev

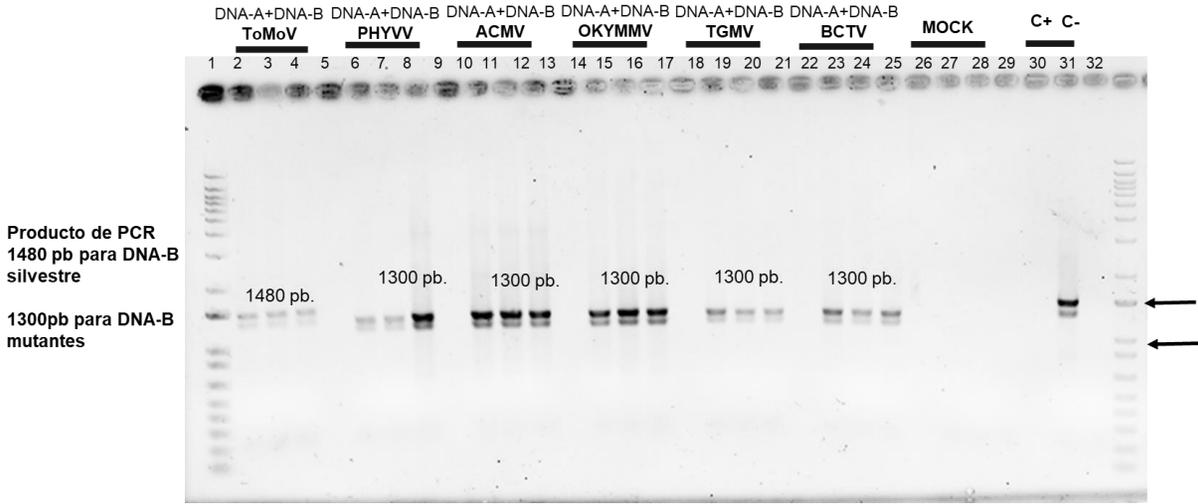


Figura 29- Productos de PCR correspondientes a los DNA-Bs replicados en plantas inoculadas con ToMoV wt y los pares de componentes mutantes examinados en este estudio. Los amplicones de los mutantes son 180 pb más cortos que el producto de PCR del componente silvestre.

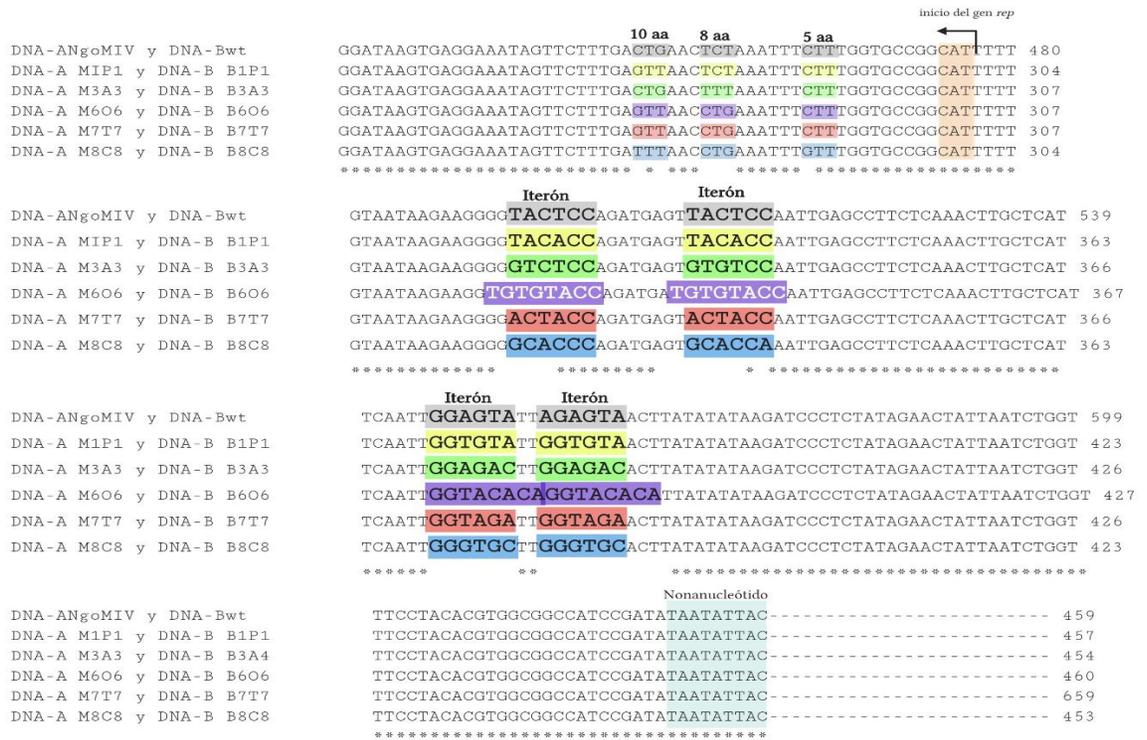


Figura 30. Alineamientos de secuencias de productos de PCR derivados de plantas infectadas que corresponden a los componentes A. Se analizaron los primeros 56 pb del inicio del gen *Rep*, en los que se localizan los elementos del IRD, así como la región intergénica donde se localizan los iterones. Se resaltan en rectángulos de colores los iterones: gris tipo ToMoV (silvestre), amarillo tipo PHYVV, verde tipo ACMV, morado tipo OkYMMV, rojo tipo TGMV, azul tipo BCTV. En el recuadro naranja se representa el ATG del gen *Rep*. El recuadro verde-grisáceo representa el nonanucleótido conservado, que incluye el sustrato endonucleolítico de la proteína *Rep*. Created with BioRender.com

De cada planta infectada se amplificó por PCR la parte relevante del componente B mutante (Figura 29), el cual es crítico para llevar a cabo una infección sistémica. Se analizaron las secuencias obtenidas por alineamientos con CLUSTAL. En la **Figura 31** se muestran los resultados obtenidos en la secuenciación de los amplicones, alineando una sección de 160 pb que corresponde a la región intergénica de DNA-B obtenidos de las plantas, y se resaltan en colores las secuencias de los iterones

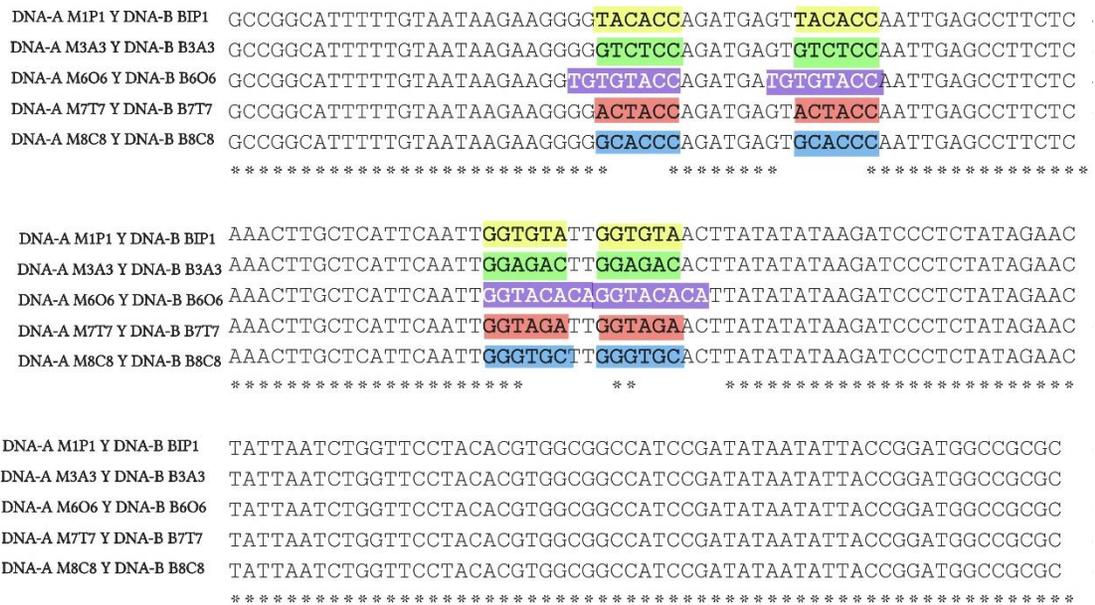


Figura 31. Alineamiento de secuencias de productos de PCR derivados de los componentes genómicos B aislados de plantas infectadas. Alineamiento de un segmento de 160 pb de los productos de secuenciación de los PCR de plantas infectadas. Se resaltan en rectángulos de colores los iterones: gris tipo ToMoV (silvestre), amarillo tipo PHYVV, verde tipo ACMV, morado tipo OkYMMV, rojo tipo TGMV, azul tipo BCTV. La secuencia del elemento conservado que forma la horquilla de replicación se encuentra en el extremo derecho de la última línea de las secuencias comparadas.

Se seleccionaron secuencias de los productos de PCR de los componentes genómicos A y B de una planta infectada con cada combinación (silvestre y mutantes) y se analizó con el software SnapGene®. En las **Figuras de la 32 a la 36** se ilustra un fragmento de la región intergénica tanto del componente A como del componente B; el DNA-A con la

presencia de los presuntos determinantes de especificidad en *cis* y en *trans* (se resaltan en cuadros de colores), además del análisis del componente B donde también se resaltan los iterones predichos. En la Figura 32 se muestra el análisis para las combinaciones DNA-A más DNA-B de **ToMoV silvestre**; en la Figura 33 se muestra el análisis de las combinaciones mutantes DNA-A **MIPI** más DNA-B **B1P1 (PHYVV)**; y la combinación DNA-A **M3A3** más DNA-B **B3A3 (ACMV)**; en la Figura 34 se presenta el análisis de la combinación DNA-A **M6O6** más DNA-B **B6O6 (OkYMMV)**, y la combinación DNA-A **M7T7** más DNA-B **B7T7 (TGMV)**; y en la Figura 35, el análisis para la combinación DNA-A **M8C8** más DNA-B **M8C8**. Como se podrá apreciar en ellas, ninguno de los componentes genómicos experimentó alteración alguna de su secuencia al replicarse en las plantas de *N. benthamiana*, de acuerdo con los datos obtenidos en el análisis de los PCRs obtenidos. En otras palabras, todos los cambios genéticos introducidos originalmente persistieron en el curso de la replicación viral, demostrando así que los mutantes generados de acuerdo con el **Modelo Predictivo** son begomovirus capaces de autoreplicación (DNA-A) y *trans*-replicación del componente B y que, por tanto, las predicciones de la correspondencia entre los DES de la proteína Rep y los iterones cognados, parecen ser esencialmente correctas. De ese modo se constata la existencia de un código combinatorio que rige las interacciones de Rep con sus secuencias de unión específica en el origen de replicación viral para iniciar la RCR.

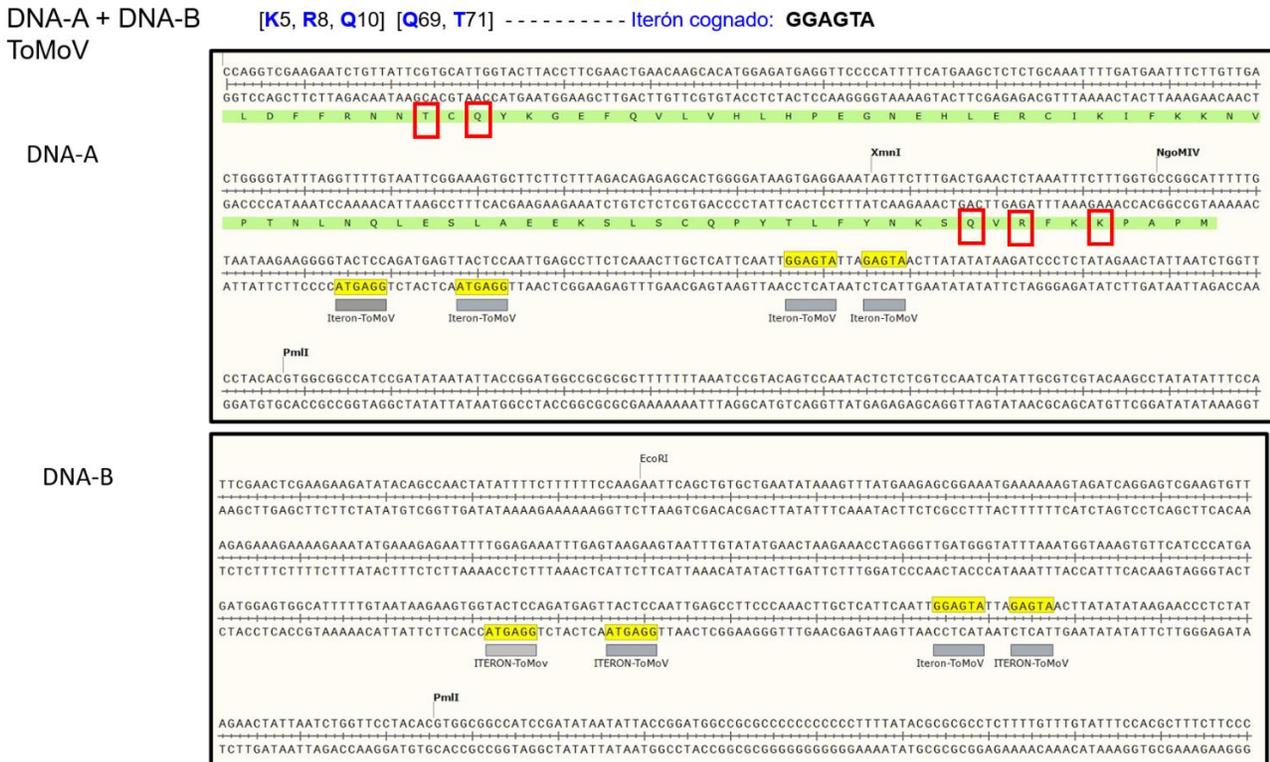


Figura 32. Análisis de secuencia con el software SnapGene®, de los PCRs recuperados de plantas infectadas con ToMoV. Se resaltan en recuadros transparentes los DEs 5,8,10, 69 y 71 y en recuadros rellenos los iterones

DNA-A M1P1+ DNA-B B1P1
TIPO PHYVV

[K5, R8, N10] [Q69, T71] ----- Iterón predicho **GGTGT**A

DNA-A



DNA-B



DNA-A M3A3+ DNA-B B3A3
TIPO ACMV

[K5, K8, Q10] [V69, T71] ----- Iterón predicho **GGAGAC**

DNA-A



DNA-B

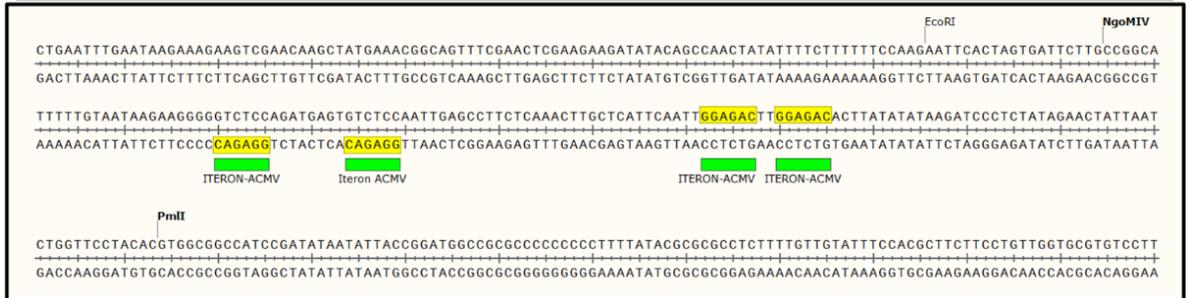
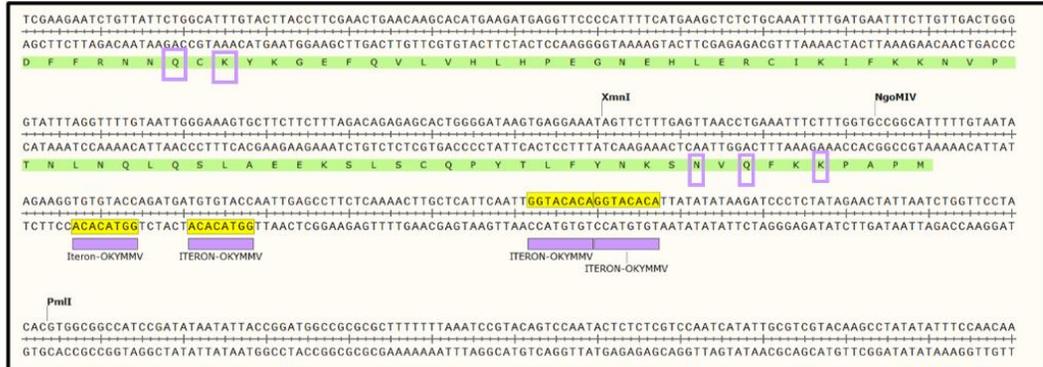


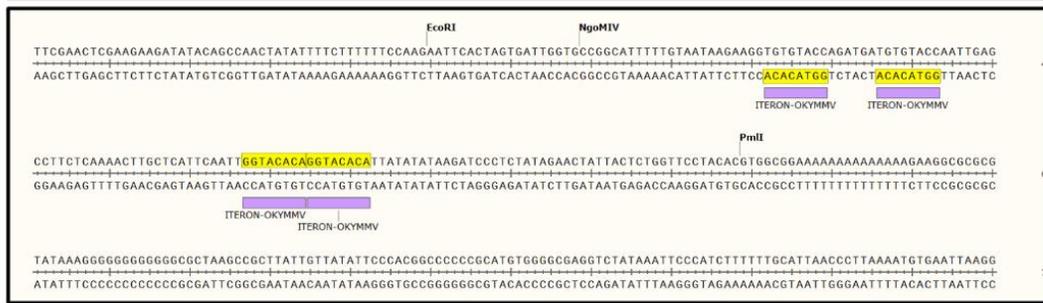
Figura 33. Análisis de secuencia con el software SnapGene®, de los PCRs recuperados de plantas infectadas con el mutante ToMoV de tipo PHYVV y ACMV. Se resaltan en recuadros transparentes los DES 5,8,10, 69 y 71 y en recuadros rellenos los iterones.

DNA-A M606+ DNA-B B606 [K5, Q8, N10] [K69, Q71] ----- Iterón predicho **GGTACACA**
 TIPO OKYMMV

DNA-A

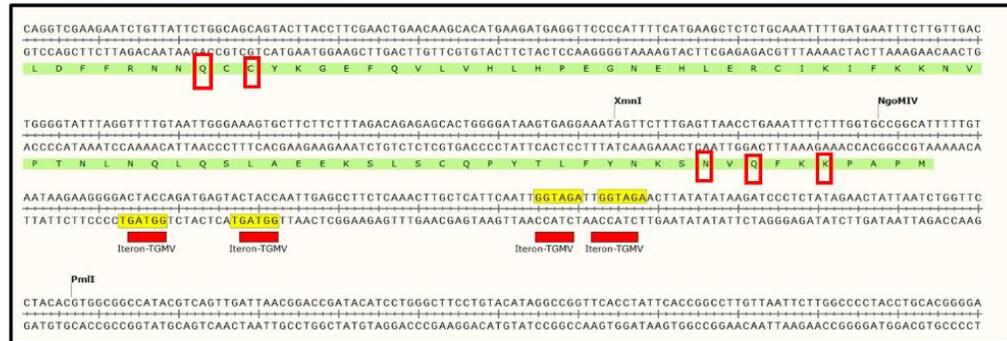


DNA-B



DNA-A M777+ DNA-B B777 [K5, Q8, N10] [C69, Q71] ----- Iterón predicho **GGTAGY**
 TIPO TGMV

DNA-A



DNA-B

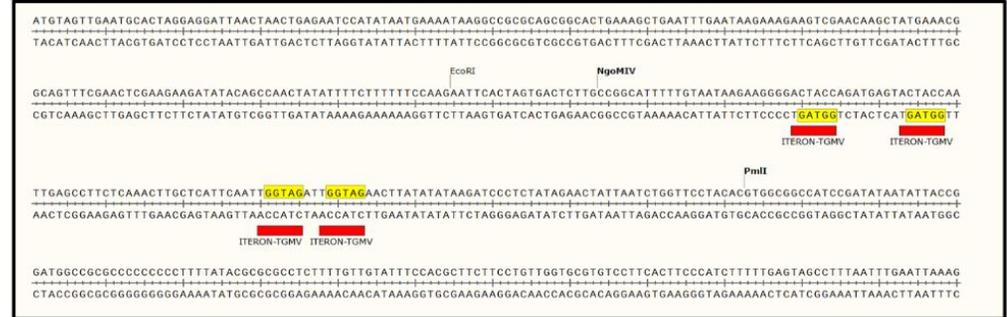


Figura 34. Análisis con el software SnapGene®, de las secuencias virales recuperadas de plantas infectadas con el mutante ToMoV de tipo OKYYMMV y TGMV. Se resaltan en recuadros transparentes los DES 5,8,10, 69 y 71 y en recuadros rellenos los iterones.

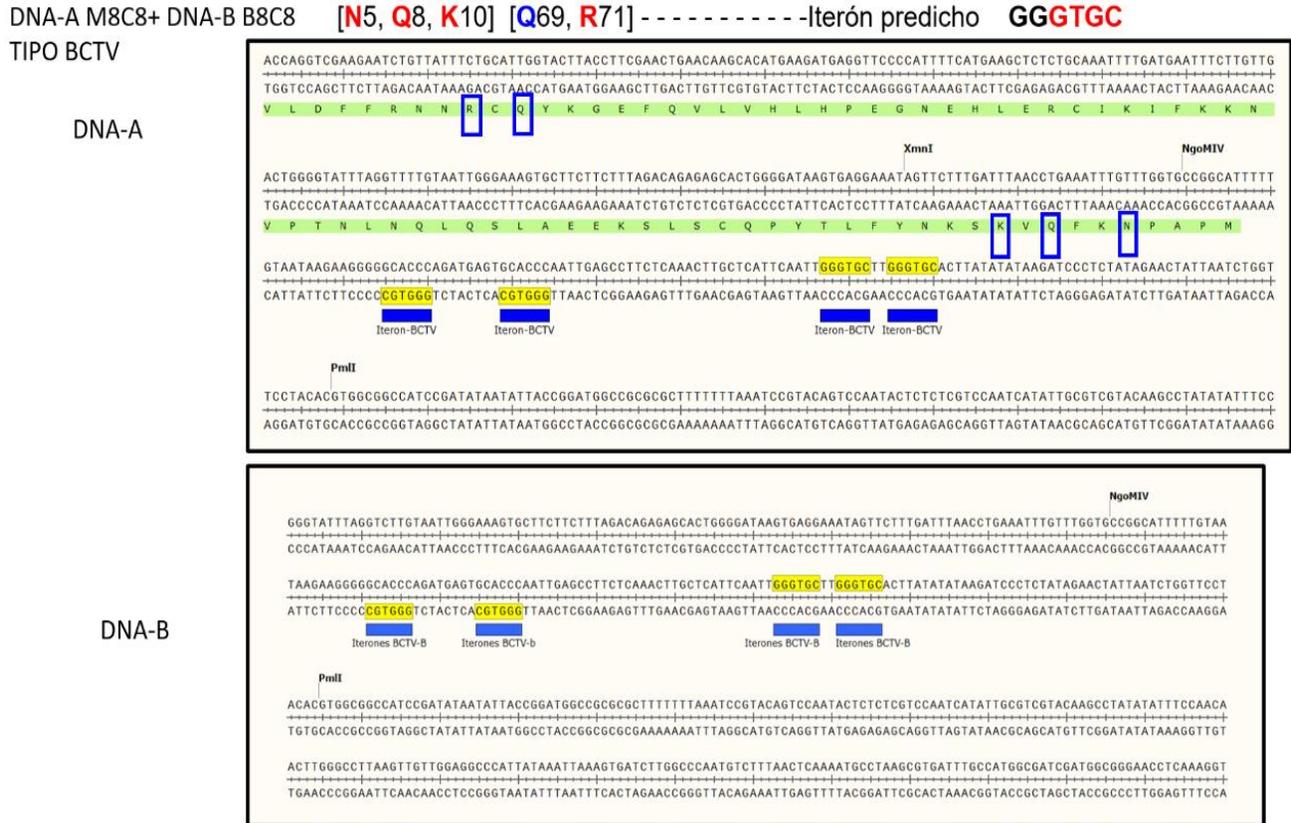


Figura 35. Análisis con el software Snap-gene de las secuencias virales recuperadas de plantas infectadas con el mutante ToMoV de tipo BCTV. Se resaltan en recuadros transparentes los DES 5,8,10, 69 y 71 y en recuadros rellenos los iterones

DISCUSIÓN

Los geminivirus constituyen un modelo idóneo para estudiar la replicación del DNA en plantas, en gran medida porque representan, junto con los nanovirus, las únicas familias de fitovirus con genomas de DNA que se reproducen en el núcleo de la célula vegetal. El resto de los fitovirus conocidos poseen genomas de RNA, o de DNA, que se replican a través de intermediarios de RNA, como los miembros de la familia *Caulimoviridae*, clasificados como pararetrovirus (Hull, 2014).

Los begomovirus representa el único subgrupo de la familia *Geminiviridae* que incluye algunas especies con un genoma segmentado (bipartita), formado por dos cadenas circulares de ssDNA muy diferentes entre sí, denominados “DNA-A” y “DNA-B”, respectivamente. Los geminivirus poseen un gen que codifica para una proteína iniciadora de la replicación por Círculo Rodante (Rep), con actividad de endonucleasa/ligasa, la cual reconoce al origen de replicación cognado (oriV) a través de su unión con alta afinidad a elementos repetidos cortos (los “iterones”), que se localizan en la vecindad de una secuencia con el potencial de formar una estructura de “horquilla” o “tallo-asa”, en cuyo ápice se localiza un nonanucleótido que es el sitio endonucleolítico de esa proteína viral (Hanley-Bowdoin et al., 2013). Rep es una proteína multifuncional que interactúa con una variedad de factores del huésped que participan en la regulación del ciclo celular, como la RBRP (“retinoblastoma-related protein”), y en procesos relacionados con la replicación del DNA en las plantas, como PCNA (“proliferating cell nuclear antigen”), RFC (“replication factor C”), y RAD54 (“recombination/ repair protein 54”). Adicionalmente, Rep interactúa con algunas quinasas como las GRIK y otras proteínas reguladoras como SCE1 (“SUMO-conjugating enzyme”), factores virales, como AC3/REn y CP (Recientemente revisados por Wu et al., 2022). Rep actúa también como un factor transcripcional negativo, inhibiendo la expresión de su propio gen al unirse a los iterones adyacente a la caja TATA (Eagle y Hanley-Bowdoin, 1994).

En trabajos previos de nuestro grupo, tanto teóricos (Argüello-Astorga et al., 1994; Argüello-Astorga y Ruiz-Medrano, 2001; Londoño et al., 2010) como experimentales (Avalos-Calleros et al., 2021; Bañuelos-Hernández, 2012; Gregorio-Jorge et al., 2010; Ramos et al., 2003) se ha establecido una serie extensa y robusta de correlaciones Rep-iterón que sugiere que la combinación de ciertos aminoácidos en 5 posiciones específicas de esa proteína viral, podrían determinar la unión preferente a una secuencia iterativa definida. Expresado de otro modo, esos residuos aminoacídicos de Rep podrían constituir los determinantes de especificidad de unión al origen de replicación cognado del virus que codifica a Rep. En el presente trabajo hemos aportado evidencia experimental crítica que apoya el modelo predictivo que se ha propuesto desde hace años (Figura 11), lo que indica que las correlaciones establecidas entre residuos aa definidos de Rep y las secuencias de sus iterones cognados, son esencialmente correctas. Sin embargo, algunas observaciones sugieren que aminoácidos en otras posiciones también modulan la afinidad relativa de Rep por los iterones cognados, al optimizar la estructura global de la proteína para su unión con alta afinidad al DNA específico

Los repetidos cortos denominados “iterones” difieren en secuencia nucleotídica aún entre virus estrechamente relacionados (ej., “cepas” o variantes de una misma especie), y pueden presentar una organización (número de copias, espaciamiento, posición relativa, etc.) diferente de acuerdo con el linaje biogeográfico al que pertenece, pero en todos los casos incluye uno o más elementos estrechamente asociados a la caja TATA del gen *Rep* (Argüello-Astorga *et al.*, 1994; Londoño *et al.*, 2010). Los primeros trabajos experimentales que mapearon el sitio de unión de *Rep* en el origen de replicación de un geminivirus, TGMV, lo ubicaron en una secuencia de 13 pb que contiene dos repetidos directos GGTAG, adyacentes a la caja TATA del gen *AC1/Rep* a los que la proteína *Rep* se unió con alta afinidad y especificidad, lo que se estableció mediante ensayos de cambio de movilidad en gel y ensayos de competencia con oligonucleótidos específicos. (Fontes *et al.*, 1994). Luego del hallazgo de una gran variedad de repetidos en el origen de replicación de los GVs conocidos en ese tiempo (Argüello-Astorga *et al.*, 1994), se utilizaron a los iterones como elementos heurísticos para identificar, tentativamente, un subdominio de las proteínas *Rep* (de 5 a 8 residuos aa), cuya estructura primaria varía entre virus que presentan diferentes iterones, pero que es similar entre virus con iterones idénticos, independientemente del huésped, origen geográfico del virus o la estructura del genoma. Este subdominio fue denominado “Dominio Relacionado al Iteron” (IRD, del inglés) (Argüello-Astorga y Ruiz-Medrano, 2001).

Posteriormente, nuestro grupo de trabajo construyó componentes A híbridos de ToMoV, para formar proteínas *Rep* quiméricas, a las cuales se les reemplazó un fragmento que codifica los primeros 11 residuos aa de la proteína por los aa equivalentes de las proteínas *Rep* de otros tres begomovirus: PHYVV, ToMoTaV, y TGMV (Bernardo-Bañuelos, 2012) La sustitución de los primeros 11 aa fueron suficientes para alterar la especificidad de unión al DNA, confiriendo a las proteínas quiméricas de ToMoV la capacidad de autorreplicar el DNA-A híbrido y de trans-replicar los componentes B heterólogos de ToMoTaV, PHYVV y TGMV, tanto en la línea celular de tabaco NT1 como en plantas de *Nicotiana benthamiana*. Inesperadamente, algunos virus híbridos mostraron una cierta capacidad para trans-replicar el DNA-B de ToMoV, estableciendo una infección con síntomas atenuados y tardíos. Esta observación sugiere que, además del IRD, podría existir otro dominio en la proteína *Rep* quimérica que, al no ser alterado, preservaría en cierto grado la capacidad de la proteína híbrida ToMoV-IRDx de reconocer y unir los iterones del DNA- B de ToMoV (Bañuelos-Hernández, 2012).

Por otra parte, en un amplio análisis heurístico-comparativo de proteínas iniciadoras de la RCR codificadas por geminivirus, Nanovirus y Circovirus, Londoño y colaboradores (2010) identificaron 5 residuos de las proteínas *Rep* geminivirales que podrían estar involucrados en determinar su preferencia por secuencias definidas. Estos cinco aminoácidos se localizan en dos mini-hebras beta ($\beta 1/\beta 5$), los residuos 5, 8, 10 en la hebra- $\beta 1$, y los residuos 69-71 en la hebra- $\beta 5$, que juntas forman una mini-lámina beta antiparalela, que se conserva en la estructura tridimensional resuelta por resonancia magnética nuclear (RMN) de las proteínas *Rep* del geminivirus TYLCSV y el circovirus porcino (Campos-Olivas *et al.*, 2002, Londoño *et al.*, 2010, Vega-Rocha *et al.*, 2007). Esta estructura forma parte de una estructura más amplia a la que pertenecen los motivos de secuencia conservados RCR-I (hebra beta-2) y RCR-II (hebra beta-4). La

estructura tridimensional de Rep revela una sorprendente similitud en la arquitectura con otras proteínas de unión a ácidos nucleicos (Campos-Olivas *et al.*, 2002).

En la primera etapa del proyecto que hemos descrito aquí, se generaron componentes genómicos A con mutaciones en el gen *Rep* de ToMoV, con base en los aminoácidos que identificamos previamente como potenciales determinantes de especificidad en *trans*. Utilizamos a ToMoV como modelo de estudio, el cual fue modificado al inicio del gen *Rep* para introducir un sitio *NgoMIV* que facilitó la mutación del extremo N-terminal de la proteína por medio de la técnica de “reemplazo de casete”. Este cambio no mostró afectar el nivel de replicación en comparación con el virus silvestre (**Figura 16**). A partir de este virus modificado se realizaron las substituciones en los codones que codifican los DEs putativos en las posiciones 5, 8, 10, y posteriormente los DEs en las posiciones 69 y 71, por las combinaciones propuestas descritas en la Figura 11, y usando diferentes técnicas de mutagénesis, que fueron descritas en la sección de Materiales y Métodos. De esa manera logramos obtener 8 DNA-A mutantes derivados de ToMoV, que se describen en la **Tabla 3**, Estos mutantes fueron examinados en ensayos de autoreplicación en protoplastos de *N. tabacum* NT1.

Los niveles de replicación de DNA de los componentes mutados en Rep mostraron diferencias muy importantes en todos los casos. En general, se observó una drástica disminución en la detección de las isoformas del DNA viral (de cadena doble y de cadena sencilla) (**Figura 16**) Esto mostró que incluso el cambio en un solo DEs, altera significativamente la interacción funcional Rep–Iterón. Los resultados obtenidos se ajustaron bien a las expectativas, de acuerdo con el modelo predictivo (**Fig.11**).

Mediante análisis *in silico*, se ha mostrado que la interacción entre la proteína Rep y el DNA se establece mediante interacciones electrostáticas y puentes de hidrógeno. Estas interacciones se forman entre los aminoácidos del dominio de unión al DNA de Rep y las bases nitrogenadas o el esqueleto de fosfatos del DNA viral (Shen *et al.*, 2018). La **figura 36** corresponde a un modelo estructural *in silico* del dominio N-terminal (1 – 120 aa) de la proteína Rep de ToMoV, así como una ampliación de la mini-lámina $\beta 1/\beta 5$ mostrando las cadenas laterales de los residuos (DEs) 5, 8, 10, 69 y 71. Adicionalmente, se muestra la estructura mini-lámina $\beta 1/\beta 5$ de las proteínas Rep de SiMSinV y OkYMMV. La comparación de las hebras $\beta 1/\beta 5$ (flechas azules y amarillas respectivamente) de los tres modelos teóricos, sugiere que las propiedades químicas de las cadenas laterales de los residuos 5, 8, 10, 69 y 71, pueden modificar la distribución de cargas electrostáticas y los patrones de puentes de hidrógeno potenciales para la interacción de Rep y el DNA viral. En este sentido, los DEs putativos de ToMoV y SiMSinV (GSK-NT) muestran los cambios en la posición 5 (K5G) de una lisina, que es un aminoácido cargado positivamente y estructuralmente grande, por un aminoácido pequeño no polar como la glicina; en la posición 8 (R8S), el cambio es de un residuo cargado positivamente a un residuo polar sin carga; en el residuo 10 (Q10K), el cambio es de una glutamina de carácter polar pero sin carga, a lisina con carga positiva; y finalmente, en la posición 69 (Q69N), el cambio de glutamina a asparagina (Q10N) aunque no representa cambios en la carga parcial, sin embargo, la presencia de un grupo CH_2 más en la glutamina puede influir en la interacción con el DNA por formación de puentes de hidrógeno. Por otro lado, comparando los DEs de ToMoV y aquellos presentes en la proteína Rep de OkYMMV, se observa que mientras

que en la posición 8 (R8Q) el cambio es de un aminoácido con carga positiva a uno polar sin carga; en la posición 10 (Q10N) ambos aminoácidos son de naturaleza polar sin carga; en el residuo 69 (Q69K), cambia de un residuo polar sin carga por uno con carga parcial positiva y con una cadena lateral más grande; y por último, el remplazo de treonina a glutamina en la posición 71 (T71Q) aunque no se modifica la carga parcial ya que ambos residuos aa no presentan carga, si poseen propiedades químicas diferentes. Lo anterior sugiere que las diferencias químicas de los residuos en el sitio de unión a ADN de Rep causan una modificación en la estructura local de la proteína. Las diferencias en la región β -1/ β 5 de las proteínas Rep proporcionan diferentes patrones de cargas electrostáticas y potenciales puentes de hidrógeno para interactuar específicamente con el DNA, lo que explicaría su mayor afinidad por iterones definidos.

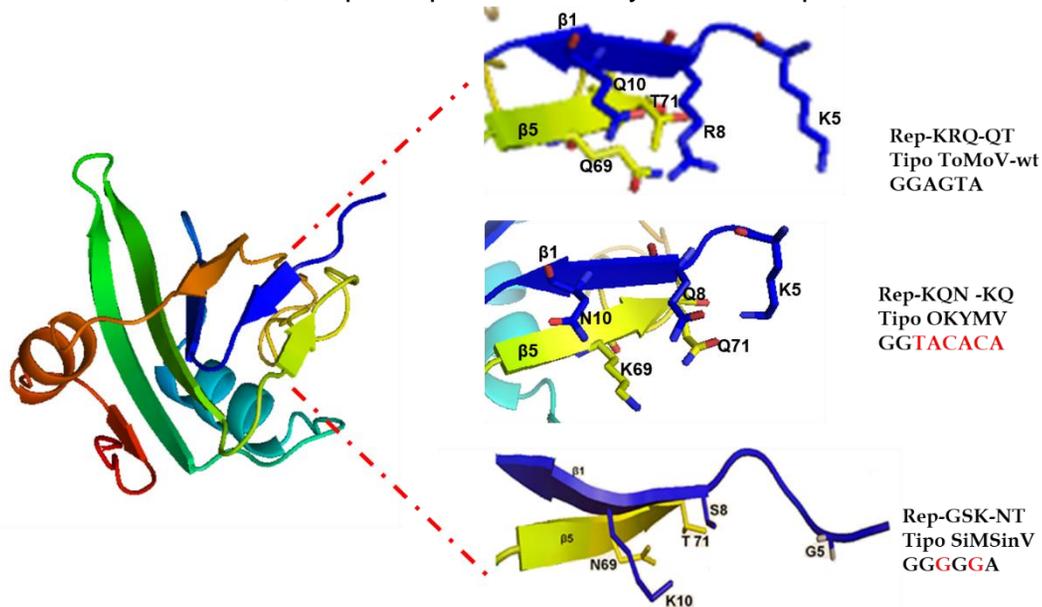


Figura 36. Modelo estructural del extremo N-terminal de la proteína Rep de ToMoV. A. Modelo por homología de la estructura terciaria del N-terminal correspondiente a los primeros 120 aa de la proteína Rep. B. Acercamiento de la mini-lámina β 1/ β 5; la hebra- β 1 representada en color azul y la hebra- β 5 representada en amarillo. Se muestran los residuos que corresponden a los DEs de las proteínas de ToMoV, SiMSinV y OkYIMV, así como cada iterón cognado, resaltando en rojo los nucleótidos que son diferentes a los presentes en ToMoV.

En las últimas etapas del estudio evaluamos la relación entre los determinantes de especificidad de las Rep mutantes y el iterón óptimo predicho para cada una de ellas. Para esto se construyeron componentes genómicos A que tienen elementos de especificidad replicativa en *cis* y *trans* presuntamente compatibles (Tabla 4). A cada DNA-A con una combinación definida de DEs se le reemplazó la región de la IR que contiene las secuencias de los iterones, reemplazando los elementos nativos de ToMoV por los iterones predichos por el modelo (Figura 19). En principio, estos nuevos DNA-As mutantes podrían haber sido examinados funcionalmente en el sistema de protoplastos de tabaco para determinar su capacidad de autorreplicación. Sin embargo, se presentaron problemas técnicos al realizar la evaluación de los mutantes dobles, pues en varios intentos consecutivos no logramos observar incluso la autorreplicación

de los controles positivos (ToMoV silvestre y ToMoV-^{NgoMIV}) que había sido muy reproducible en los primeros ensayos. Después de varios intentos, llegamos a la conclusión que la línea celular NT-1 utilizada había experimentado cambios fisiológicos y/o epigenéticos, y no era ya útil para realizar los ensayos de autorreplicación. Al no encontrar una forma de obtener la línea celular NT-1 original o una nueva fuente de dichas células, buscamos un sistema alternativo para los ensayos de autorreplicación transitoria

La técnica elegida fue el ensayo de autoreplicación en discos de hoja de *N. benthamiana* (Fig. 21). Este tipo de ensayo requiere que los componentes genómicos virales se encuentren en forma dimérica o hemidimérica, y sean clonados en un vector binario, que ha de ser transformado en una cepa específica de *Agrobacterium tumefaciens*, la cual se utilizará para infiltrar hojas de *N. benthamiana*. Por esta razón decidimos seleccionar solo cinco DNA-As mutantes representativos de la diversidad y el grado de robustez de las correlaciones Rep/ Iteron de nuestro modelo teórico. Las mutantes seleccionadas fueron las del tipo PHYVV, ACMV, OkYMMV, TGMV y BCTV (Tabla 5). Después de varias pruebas para poner a punto la técnica, logramos observar la autorreplicación de los DNA-As mutantes (Fig.22). Estos resultados, indican que las proteínas Rep mutantes reconocieron los iterones en el origen de replicación viral mutante, y por lo tanto, que la replicación por círculo rodante fue posible, de modo similar al virus silvestre.

Experimentos de agroinoculación de componentes mutantes en plantas de *Nicotiana benthamiana*.

Aunque los datos obtenidos hasta este punto apoyaban la noción de que las proteínas Rep con DEs alterados pueden unirse a los iterones predichos por el modelo teórico, nos interesó determinar si las proteínas Rep mutantes pueden reconocer y unir a los iterones predichos cuando estos se encuentran en el componente B de ToMoV, y si pueden inducir así una infección sistémica en plantas de *N. benthamiana*. Para ello, se construyeron componentes B mutantes a partir del componente B de ToMoV, en el que se substituyó un segmento *EcoRI /PmlI* de la región intergénica, donde se ubican los iterones de ToMoV-B (Figura 16), con la región análoga, aunque más corta *EcoRI/PmlI* de los componentes A mutantes, con sus iterones cognados (Figura 16). Dado que en el DNA-B wt de ToMoV la región intergénica limitada por los sitios de restricción mencionados es más extensa que en los DNA-As mutantes, los DNA-B mutantes tienen una región intergénica más corta. Los DNA-Bs mutantes se convirtieron a su forma infectiva (hemidimérica) y se transformaron en *Agrobacterium tumefaciens*. El primer ensayo de co-infección fue inocular a los componentes A dobles mutantes en presencia del DNA-B de ToMoV silvestre para establecer si podrían presentar un fenómeno de pseudo-recombinación. Ambos componentes se agroinfiltraron (proporción 1:1) en plantas de *N. benthamiana*, sin embargo, las plantas no mostraron síntomas (datos no mostrados) y al establecer la presencia del DNA-B por PCR, no hubo amplificación a los días 14 y 21 post inoculación (**Figura 25**), lo que indica que los componentes A mutantes no fueron capaces de reconocer y unir a los iterones presentes en el componente B silvestre.

Posteriormente, se co-inocularon ambos componentes A y B mutantes en plantas de *N. benthamiana*, y después de 14 y 21 dpi se tomaron muestras de hojas nuevas (apicales) de cada combinación, y se tomó evidencia fotográfica del desarrollo de síntomas en las plantas infectadas a los tiempos mencionados.

Las plantas infectadas con los componentes silvestres de ToMoV presentaron síntomas visibles partir de los 10 dpi, aunque se hicieron más evidentes a los 14 y 21 dpi, presentando síntomas como amarillamiento de las hojas, con cierto grado de moteado, rizado leve de hojas y retraso en el crecimiento, lo que indicaba que nuestro sistema de infección por agroinfiltración, usando clonas infectivas, funcionaba adecuadamente.

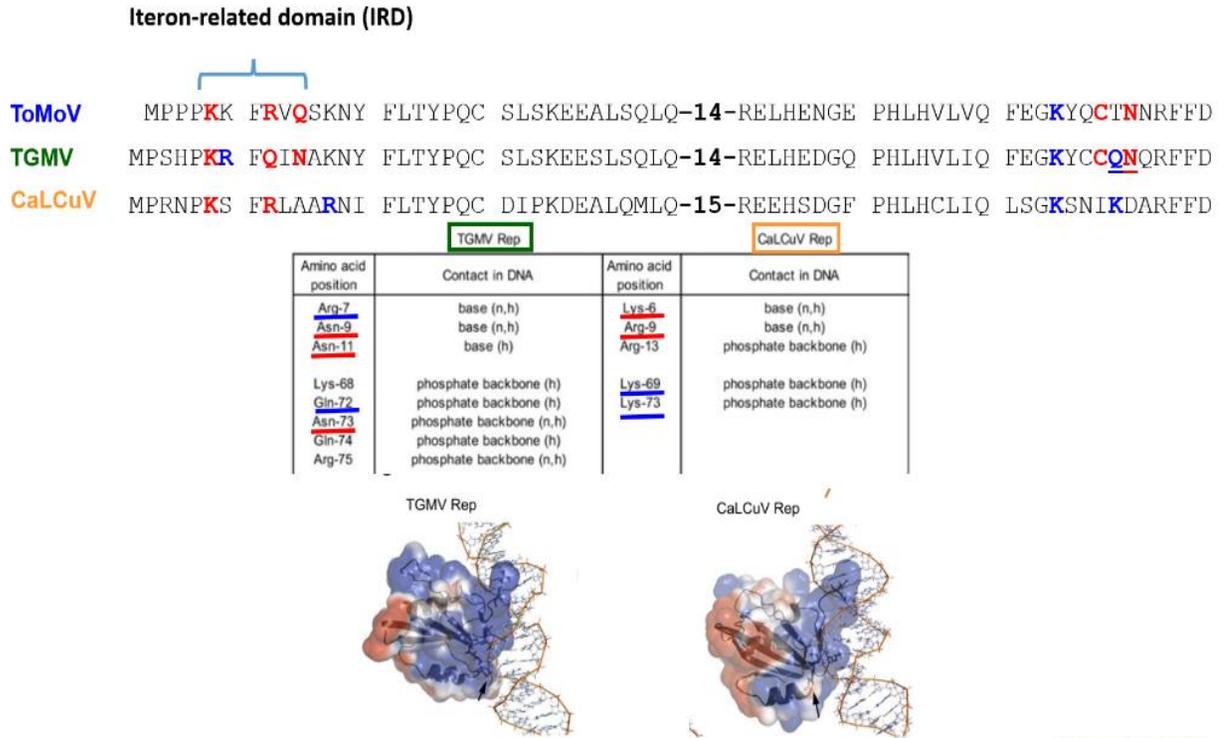
En la valoración de la sintomatología de las plantas infectadas con los componentes A y B mutantes, los síntomas se evidenciaron a partir del día 14 post-inoculación. Dado que los síntomas no fueron idénticos entre los virus mutantes y el virus silvestre, lo que inferimos es que las mutaciones en los codones 69 y 71 del gen *Rep*, que se traslapan con el gen AC4, probablemente provocaron cambios en este último gen y la proteína codificada, que podrían haber alterado el fenotipo infectivo. La proteína AC4 de begomovirus está involucrada en múltiples funciones, como la inducción de síntomas, hiperplasia y la supresión del silenciamiento génico, entre otros (Hanley-Bowdoin, et al., 2013; Zeng, et al 2018; Fondong et al., 2019). Se ha demostrado que mutaciones en el gen C4 reducen la acumulación del virus en plantas (Ismayil et al., 2018).

Se tomaron muestras de las hojas apicales de las plantas infectadas y se procesaron para obtener DNA total, que se usó como templado para la detección por medio de PCR de ambos componentes virales mutados. Los productos de PCR se secuenciaron, y los resultados, se analizaron usando el Software SnapGene®. (Figuras 32-35). El análisis evidenció que cada uno de los productos de PCR obtenidos tiene intactas las secuencias que corresponden a los hipotéticos DES y a los iterones de cada uno de los componentes A y B mutantes, es decir, que las mutaciones introducidas no sufrieron reversiones y que las secuencias de los componentes virales corresponden a los DNAs con los que se realizó la agroinfiltración. Esto demostró que los DNA-A y DNA-B mutantes son capaces de inducir una infección sistémica en plantas, y que por lo tanto las proteínas Rep modificadas fueron capaces de reconocer los iterones en el origen de replicación mutante de ambos componentes virales, con lo cual se confirma, en principio, la validez del modelo hipotético general ilustrado en la Fig. 11. Estos hallazgos constituyen una importante contribución al conocimiento del código de reconocimiento de los iterones por la proteína Rep.

Datos de otros estudios que apoyan las conclusiones de este trabajo.

En el curso del desarrollo del presente estudio se publicaron algunos trabajos que son coherentes con nuestros resultados. Por ejemplo, Shen et al. (2018) llevaron a cabo un modelado *in silico* de las interacciones de las proteínas Rep (aa 1-120) de dos begomovirus muy diferentes, TGMV y cabbage leaf curl virus (CaLCuV), con el DNA viral correspondiente. En sus resultados, reproducidos parcialmente en la **Fig. 37**, ellos identificaron a los residuos aminoácidos Arg7, Gln9 y Asn11 de la proteína Rep de

TGMV (equivalentes a los residuos 6, 8 y 10 de ToMoV) y los residuos Lys6 y Arg9 de la proteína Rep de CaLCuV (equivalentes a los residuos 5 y 8 de ToMoV).



Shen et al. 2018

Figura 37. Modelo estructural *in silico* de la interacción proteína Rep-DNA de TGMV and CaLCuV. Los aa que podrían funcionar como DEs están representados en rojo y azul, que son equivalentes a algunos residuos en nuestro modelo de ToMoV. Modificado de Shen *et al.*, 2018.

Más recientemente, se publicó un estudio en el que se describieron dos cepas del virus RhGMSV (*Rhynchosia golden mosaic Sinaloa virus*), una aislada en Guasave, Sinaloa, y otra en Colima (RhGMSV-GU y RhGMSV-CO). Sus iterones son marcadamente diferentes, TGGAGGAY y TGGAGTCC, respectivamente. Sus proteínas Rep son idénticas en los primeros 24 aa, que incluyen la totalidad del IRD, y el alineamiento de los primeros 130 aa, donde reside el dominio de unión al DNA de las proteínas Rep, mostró que existen solo dos aminoácidos diferenciales entre las dos cepas, E69K y D71N. Este último residuo aa es parte de la hebra β -5 de Rep. En experimentos para determinar la capacidad de formar pseudo-recombinantes entre las dos cepas de RhGMSV, se confirmó que éstas no fueron compatibles en replicación. Sin embargo, en un experimento de pseudo-recombinación entre el DNA-A de RhGMSV-CO con el DNA-B de Pepper golden mosaic virus (PepGMV), un begomovirus muy divergente,

pero con iterones similares, se pudo demostrar la formación de un pseudo-recombinante viable, que produjo una infección atenuada pero sistémica en *N. benthamiana* (Avalos-Calleros et al., 2021).

Las implicaciones de nuestro estudio con begomovirus quiméricos derivados de ToMoV pueden ser muy amplias, no solo por extensión a otros géneros de los *Geminiviridae*, sino porque Londoño et al. (2010) mostraron que miembros de varias familias de virus, con genomas circulares de cadena sencilla, como *Circoviridae* y *Nanoviridae/Alphasatellidae*, codifican proteínas *Rep* que incluyen hebras beta muy cortas que potencialmente conforman mini-láminas beta antiparalelas cuya estructura primaria correlaciona con los iterones que constituyen sus potenciales sitios de unión en el origen de replicación viral. Las familias antes mencionadas han sido recientemente agrupadas en un nuevo phylum, *Cressdnaviricota*, que aglutina 7 familias de virus con genomas circulares de ssDNA (Krupovic et al., 2021). Es probable, por tanto, que más familias virales codifiquen proteínas *Rep* análogas a las de los geminivirus, y que el código de reconocimiento *Rep/iterón* de cada una de ellas pueda ser descifrado usando la metodología heurística-comparativa que utilizamos con los geminivirus.

Capítulo 2

Explorando el origen del ancestro no-begomoviral del linaje del *Squash Leaf curl virus*

El estudio de la evolución de los begomovirus, el género más diversificado y de más amplia distribución geográfica de la familia *Geminiviridae*, ha sido objeto de múltiples análisis desde hace casi tres décadas, desde el estudio pionero de Rybicki (1994). El número de especies geminivirales descritas ha crecido exponencialmente en la última década, en gran medida debido al advenimiento de las tecnologías de Secuenciación de Nueva Generación (NGS, del inglés), que ha hecho posible el ensamblaje de genomas virales a partir de muestras ambientales y de plantas, lo que ha resultado en la adición de varios géneros a la familia *Geminiviridae* en los últimos años (Zerbini et al., 2017; Fiallo-Olive, 2021). En el caso de los begomovirus (BGVs), la tradicional distinción de BGVs del Viejo y del Nuevo Mundo ha dado paso a la definición de linajes que comparten ciertas características moleculares y una historia evolutiva común, que permiten su distinción en subclados mejor definidos y relevantes, lo que enriquece el conocimiento de la diversidad de estos subgrupos geminivirales.

En su tesis doctoral, Iliana Torres-Herrera (2015), se enfocó en el estudio de un grupo atípico de begomovirus (BGVs) del Nuevo Mundo, el llamado “Clado del Squash leaf curl virus (SLCV)”, que se distingue del resto de los BGVs del continente americano por la peculiar organización de sus iterones en el origen de replicación, en el que dos iterones extras flanquean al elemento tallo-asa (“horquilla conservada”), además de desplegar otros 3 o 4 elementos repetidos en la configuración típica de los BGVs del NM (Argüello-Astorga, 1994). Adicionalmente, estos virus codifican una proteína Rep que es muy divergente en sus primeros 160 aminoácidos de la proteína homóloga del resto de los begomovirus, y poseen un gen que se traslapa completamente con el gen *Rep*, que codifica una proteína que denominamos AC4*, para distinguirla de las proteínas C4/AC4 de otros begomovirus “típicos”, debido a su evidente divergencia en secuencia de aminoácidos, que sugiere que no es una proteína homóloga de éstas últimas.

Tomando en consideración el tema central de nuestro proyecto de tesis, esto es, la indagación de los residuos aminoácidos de Rep que determinan su alta afinidad por los iterones en el origen de replicación cognado, nos pareció de gran importancia investigar las diferencias en la estructura y origen evolutivo de las proteínas Rep de los virus del clado SLCV, las cuales reconocen iterones de 8 nucleótidos con un consenso diferente (TGGWGD^{MC}) a los de otros begomovirus (GGNRNN) que suelen presentar iterones de 6 nucleótidos, aunque existen algunos más extensos. En esta parte de mi trabajo de tesis orienté mis esfuerzos a tratar de responder algunas de las preguntas

que no pudieron contestarse en el estudio previo. Dado que el trabajo que llevé a cabo al abordar este segundo tema, quedó incorporado en un artículo en la revista *Viruses* (de acceso libre) que incluyó los hallazgos no publicados de la tesis doctoral de Torres-Herrera (2015) y los míos propios, en lo que sigue describiré tan solo los aspectos metodológicos y los hallazgos más relevantes de lo que realice por mi parte, como continuación del trabajo de Torres-Herrera, y con el apoyo de mis asesores. En la primera fase de este nuevo trabajo, llevamos a cabo la búsqueda de miembros del linaje del SLCV, y nuevos geminivirus que fueron descritos del 2015 en adelante, que no fueron incluidos en el trabajo previo, por razones obvias. Se generaron filogenias muy amplias basadas en los componentes genómicos “A” completos de los BGVs bipartitas, o el genoma único de los virus monopartitas (datos no mostrados), y posteriormente se seleccionaron representantes destacados de los principales linajes virales discernibles. Posteriormente se realizaron colecciones de proteínas Rep truncadas en el aminoácido 162, que identificamos como el punto de divergencia más robusto de los diversos alineamientos de proteínas Rep de los miembros del clado del SLCV con otros linajes begomovirales. Como se puede apreciar en la **Figura 2.1**, la filogenia basada en la comparación de los primeros 162 residuos aa de la proteína Rep de los diversos linajes comparados, muestra claramente dos grandes grupos, uno de los cuales comprende a los BGVs del clado del SLCV, y dos curtovirus atípicos, Horseradish curly top virus (HCTV) y Spinach severe curly top virus (SpiSCTV); por otra parte, el segundo grupo, muchísimo más extenso en términos filogenéticos, incluye a los BGVs típicos tanto del NM como del VM, pero también algunos miembros de los géneros *Curtovirus*, *Maldovirus*, *Topocuvirus*, y *Turncurtovirus*. En este segundo grupo se ubican también algunas especies no clasificadas todavía en géneros específicos, y algunos geminivirus que son conocidos solamente por secuencias que existen como fósiles moleculares en los genomas de ciertas plantas, y que son denominadas “Secuencias Virales Endógenas” (EVS, del inglés) las cuales analicé a partir de la literatura existente o por búsqueda dirigida en la base de datos del NCBI-GenBank (indicadas en la Fig. 2.1 por un símbolo especial).

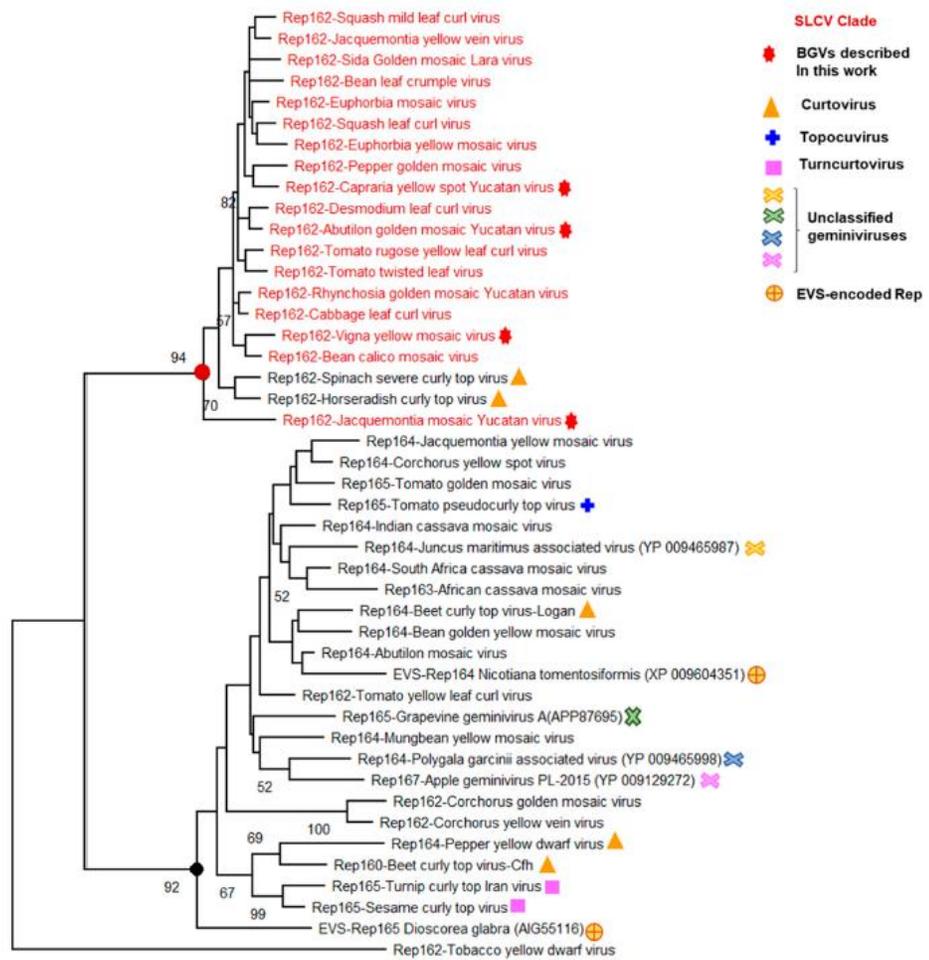


Figure 2.1. Relación filogenética basada en la comparación del dominio amino-terminal (1-162) de miembros del clado del SLCV (letras en rojo) y sus homólogos en otros subgrupos de geminivirus. El árbol filogenético fue inferido por el método de Máxima Verosimilitud basado en el modelo general de la transcriptasa reversa. El porcentaje de árboles en los que los taxones están agrupados se muestra próximos a las ramas (1000 iteraciones). El círculo rojo indica el nodo del clado del SLCV, en tanto que el círculo negro indica la rama principal del nodo de proteínas Rep 162 de otros geminivirus, incluyendo algunos EVS (secuencias virales endógenas).

Para nuestros fines, nos pareció relevante dar especial importancia a la búsqueda de EVS, porque es un conjunto de secuencias presentes en los genomas de las plantas, que no han sido utilizados ampliamente para establecer la relación filogenética con geminivirus existentes. En el caso de los miembros del linaje del SLCV, cuyo ancestro no begomoviral no ha sido identificado, proceder con ese tipo de búsqueda en las bases de datos genómicos de plantas, resulta especialmente prometedor en el caso de que no se encuentre tal ancestro entre los geminivirus descubiertos por técnicas convencionales o metagenómicas. Lo que concluimos de nuestro muy amplio análisis comparativo es que, a la fecha, no se han reportado EVS

que presenten una clara relación con virus del clado del SLCV. Sin embargo, en el intento por descubrir en las bases de datos publicas, virus con similitudes significativas a los begomovirus mencionados, decidimos usar las secuencias de los primeros 120 residuos (dominio catalítico) de Rep codificadas por virus específicos como SLCV, ToSLCV o el recientemente descrito Cucumber chlorotic leaf curl virus (CuCLCV) (Sánchez-Chavez et al., 2020), como secuencia problema (“query”) para interrogar por secuencias similares en las bases de datos públicas, excluyendo naturalmente a los geminivirus. De esa manera encontramos secuencias virales endógenas (EVS) en *Coffea arabica* que codifican proteínas (XP_027067771) con 51% de identidad y 73% de similitud (**Figura 2.2a**), o una proteína de replicación de un virus circular de Persimon (**Figura 2.2b**).

A

uncharacterized protein LOC113693425 [Coffea arabica]
Sequence ID: [XP_027067771.1](#) Length: 272 Number of Matches: 1

Score	Expect	Method	Identities	Positives	Gaps
137 bits(346)	2e-37	Compositional matrix adjust.	61/119(51%)	88/119(73%)	1/119
<u>Query_1</u>	MPRNPFLFRLAAKNIFLTYPQCDIPKDEAIEMLQNLPWSVVKPTYIRVAREEHADGFPHL 60				
	MPR P+ F + AKNIFLTYP+C +PK +A++ ++NL + + P YIRVA+E H DG PHL				
<u>Sbjct_1</u>	MPRAPSSFFINAKNIFLTYPKQALDAIRNLQFPI-SPIYIRVAKETHEDGSPHL 59				
<u>Query_61</u>	HCLIQLSGKSNIKDARFFDLTHPRRSANFHPNAQAAKDTNAVQNYITKEGDYCESGQYK 119				
	HCLIQ GK + ARFFD+ P ++ FHPN Q A++++AV++YI+K GD+ E G+++				
<u>Sbjct_60</u>	HCLIQFEGKFRTESARFFDIKSPISNSMFHPNVQGARNSSAVRDYISKYGDFVEWGEFR 118				

B

replication associated protein [Lake Sarah-associated circular virus-6]
 Sequence ID: [YP_009237504.1](#) Length: 317 Number of Matches: 1

Score	Expect	Method	Identities	Positives	Gap
92.4 bits(228)	2e-19	Compositional matrix adjust.	49/110(45%)	65/110(59%)	6/1
<u>Query_8</u>					
<u>Sbjct_24</u>					
<u>Query_68</u>					
<u>Sbjct_80</u>					

```

FRLAAKNIFLTYPQCDIPKDEAIEMLQNLPSVVKPTYIRVAREEHADGFPHLHCLIQLS 67
FR_AKN LT+PQCD+ K++A + + + KP Y+ VA EEH DG PHLH LI S
FRATAKNFALTFPQCDMNKEDAFQKIMDEH----KPAYLCVAEEEHKDGSPHLHALISFS 79

GKSNIKDARFFDLTHPRRSANFHPNAQAAKDTNAVQNYITKEGDYCESGQ 117
K NIKD + FDL+ + FH N Q+ K + YI K+G Y E+ +
SKKNIRDKPHFDLSD--NESTFHCNIQSCKSVADWKRYIKKDGKRYKETEE 127

```

Figura 2.3- Alineamiento del dominio endonucleolítico de la proteína Rep de Tomato severe leaf curl virus (ToSLCV) y un virus aislado de una comunidad de invertebrados en el fondo del Lago Sarah, de Nueva Zelanda.

Estas observaciones no indican, por supuesto, que un virus animal se encuentre en la base o en una rama del árbol filogenético al que pertenece el ancestro no-begomoviral de los miembros del clado del SLCV. Más bien, lo que sugieren es que algunas proteínas Rep pertenecientes a la superfamilia de las endonucleasas HUH (Chandler et al., 2013) pueden haber convergido en las propiedades de su dominio de unión al DNA para reconocer iterones idénticos, lo cual plantea la posibilidad de que el ancestro del clado del SLCV haya adquirido el módulo de replicación atípico que lo caracteriza, a partir de un evento de recombinación entre un begomovirus del NM muy antiguo, con un virus con genoma circular de ssDNA que infectaba plantas, análogo al virus circular del persimón mencionado en la Fig.2b.

Análisis comparativo de la proteína AC4* de los virus del clado SLCV.

La proteína AC4 de los begomovirus es una proteína multifuncional que se considera un determinante de patogenicidad, un supresor del silenciamiento génico, y un factor que promueve la progresión del ciclo celular en ciertos casos bien documentados (revisado por Fondong, 2018). Una parte importante de las propiedades de esa proteína se asocian a la capacidad de ser miristilada, para anclarse a la membrana celular, y por ello presenta un motivo de miristilación, MGXLIS, en el que la glicina en la posición 2 juega un papel crítico. El tamaño promedio de la proteína C4/AC4 de begomovirus y otros geminivirus es de 85aa. Sin embargo, la proteína AC4* de los

miembros del linaje del SLCV es mucho más larga, cercana a los 120 aa, y carece del motivo de miristilación presente en sus presuntos homólogos. El alineamiento de secuencia de esa proteína con las C4/AC4 de otros geminivirus muestra que la identidad de secuencia polipeptídica es demasiado baja (<23%) para considerarse proteínas homólogas. Consecuentemente, la función de AC4* de los virus del clado SLCV es, muy probablemente, bastante diferente a las de otras proteínas geminivirales con ese nombre.

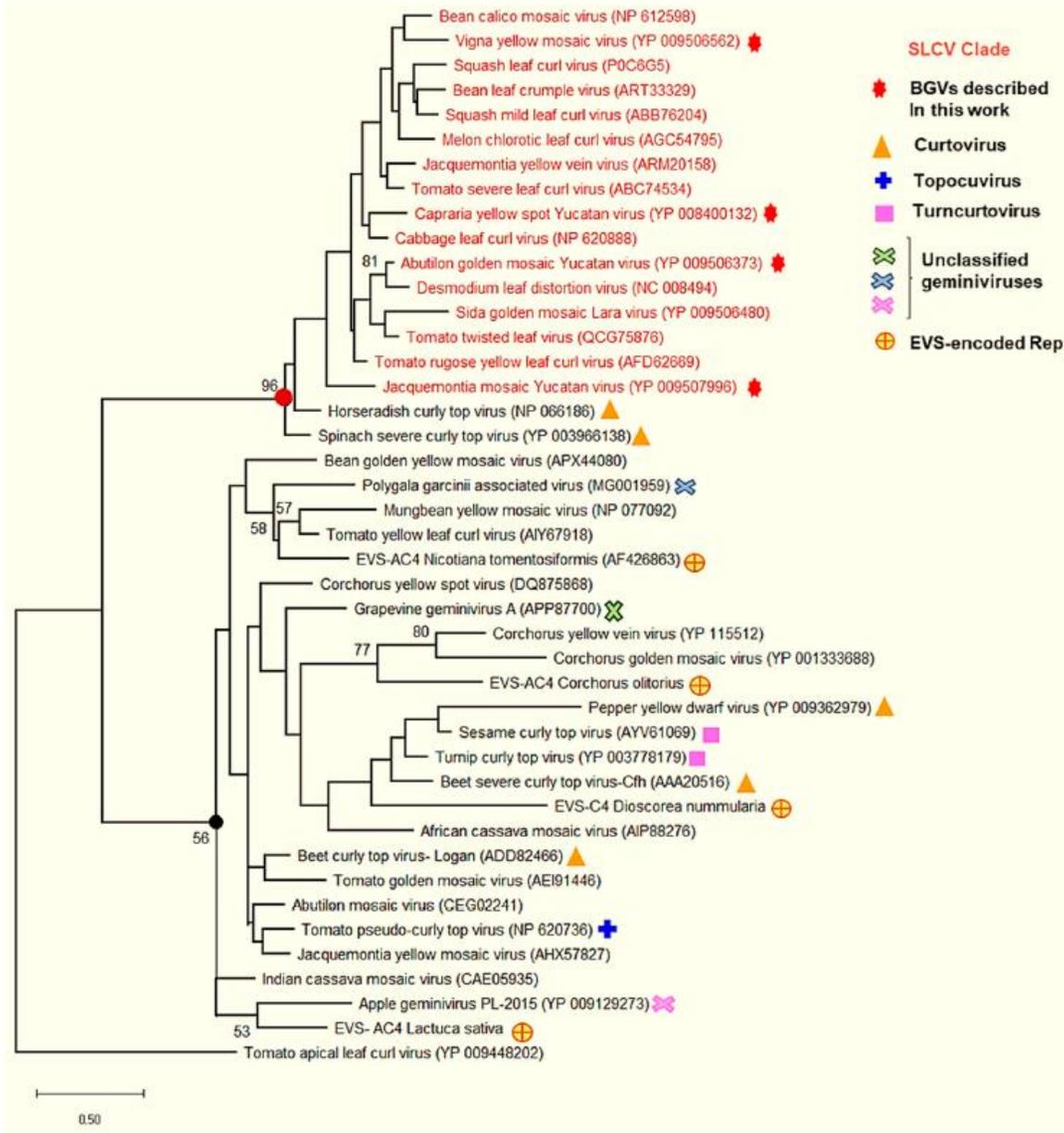


Figura 2.4.- Relación filogenética entre las proteínas AC4* codificadas por los virus del linaje del SLCV (letras en rojo) y las proteínas C4/AC4 de otros linajes de geminivirus. El árbol filogenético fue inferido mediante el método de Máxima Verosimilitud basada en el modelo basado en la matriz JTT. El porcentaje de árboles en los que los taxones asociados se agruparon juntos se muestra cerca de las ramas (1000 iteraciones). Reproducido de Torres-Herrera *et al.*, 2019.

En la **Figura 2.5** se muestra un alineamiento de varias proteínas AC4* con proteínas C4/AC4 de otros geminivirus, en el que puede apreciarse que en las primeras no existe

el motivo de miristilación que se encuentra prácticamente en todas las proteínas C4/AC4, con la excepción de algunas EVS.



Figure 2.5. Alineamiento del dominio N-terminal de la proteína AC4* de los virus del linaje del SLCV y dos curtovirus recombinantes, con el dominio análogo de las proteínas C4/AC4 codificadas por algunos begomovirus del NM y el VM, otros miembros del género *Curtovirus*, *Topocuvirus*, algunos geminivirus no asignados a un género específico, así como cinco EVS (Secuencias Virales Endógenas). Los residuos aminoácidos idénticos o similares están sombreados en amarillo o en azul para enfatizar las semejanzas entre las secuencias alineadas. El residuo de Glicina del motivo de miristilación está marcado en rojo. (reproducido de Torres-Herrera et al., 2019)

El nombre de los virus y los EVS, de la Fig. 2.5, así como sus números de acceso, se enlistan en la **Tabla 2.1**.

Tabla 2.1 – Nombre y # de acceso de geminivirus y secuencias virales endógenas (EVS) del alineamiento de proteínas AC4*, C4 y AC4 de la Fig. 2.5.

<i>Virus</i>	Acronimo	No, de acceso- GenBank
<i>Abutilon golden mosaic Yucatan virus</i>	AbGMYV	KC430935
<i>Cabbage leaf curl virus</i>	CbLCuV	MH248136
<i>Bean calico mosaic virus</i>	BCaMV	AF110189
<i>Tomato common mosaic virus</i>	ToCoMV	NC_010835
<i>Sida yellow leaf curl virus</i>	SiYLCV	EU710750
<i>Tomato rugose yellow leaf curl virus</i>	ToRYLCV	JN381815
Curtovirus-tipo II		
<i>Horseradish curly top virus</i>	HrCTV	U49907
<i>Spinach severe curly top virus</i>	SpSCTV	GU734126
Begomovirus del Nuevo y el Viejo Mundo		
<i>Tomato golden mosaic virus</i>	TGMV	NC_001507
<i>Rhynchosia golden mosaic virus</i>	RhGMV	NC_010294
<i>Tomato yellow leaf curl Vietnam virus</i>	TYLCVV	EU189150
<i>Mimosa yellow leaf curl virus</i>	MiLCV	DQ641695
<i>East African cassava mosaic Zanzibar virus</i>	EACMZV	MK059410
<i>Cassava mosaic Madagascar virus</i>	CMMdV	HE617299
Curtovirus- tipo I		
<i>Beet curly top virus - California [Logan]</i>	BCTV-Cal	NC_001412
<i>Beet curly top virus- Idaho</i>	BCTV-Idaho	KX867057
Topocuviruses		
<i>Tomato pseudo-curly</i>	TPCTV	X84735

top

Geminivirus no clasificados

<i>Grapevine geminivirus A</i>	GraGV-A	KX618694
<i>Apple geminivirus</i>	AGmV	KM386645
<i>Juncus maritimus-associated virus</i>	JmaV	MG001958
<i>Polygala garcinii-associated virus</i>	PgaV	MG001959

Secuencias Virales Exógenas (EVS)

<i>Lactuca sativa cultivar Salinas unplaced genomic scaffold</i>	EVS- <i>Lactuca sativa</i>	NW_019676544
<i>Corchorus olitorius cultivar JRO-524</i>	EVS- <i>Corchorus olitorius</i>	LLWS00000000
<i>Nicotiana tomentosiformis unplaced genomic scaffold</i>	EVS- <i>Nicotiana tomentosiformis</i>	NW_008941319
<i>Dioscorea nummularia endogenous virus EGV2</i>	EVS- <i>Dioscorea nummularia</i>	KJ629225
<i>Coffea arabica geminivirus-like replication protein</i>	EVS- <i>Coffea canephora</i>	XM_027241759

Hallazgo de un putativo elemento *cis*-regulador de la transcripción asociado al Origen de Replicación en los virus del linaje del SLCV.

Una característica de la región intergénica que distingue a los virus del clado SLCV de otros BGVs del NM, además del número, arreglo y secuencia de sus iterones, es que carecen de una G-box, que es unida por factores transcripcionales de la planta y es crítico para la actividad del promotor *Rep* (Orozco et al., 1998). En vez de ese elemento, estrechamente asociado al extremo 5' del elemento capaz de formar la estructura tallo-asa esencial para que *Rep* pueda cortar la cadena+ del DNA viral, se encuentra en la misma posición relativa, una secuencia altamente conservada, GGGGCAAAA. Una búsqueda en bases de datos especializados en elementos *cis*-reguladores de genes de plantas, como PLACE (Higo et al., 1999) y SoftBerry (<http://www.softberry.com/>) identificaron esa secuencia como sitio de unión del factor E2F de *Arabidopsis thaliana* y *N. tabacum*. Muy importante es el hecho de que ese factor transcripcional es parte de la maquinaria molecular responsable de la progresión del ciclo celular en las células vegetales (Chaboute et al., 2000; Vandepoele et al., 2005). De modo interesante, la posición del sitio de unión putativo E2F en la vecindad del elemento de horquilla conservado en el origen de replicación de geminivirus, es

consistente con lo observado en muchos otros geminivirus, en los que un sitio de unión a un factor transcripcional se localiza inmediatamente adyacente al elemento tallo-asa del origen de replicación (Argüello-Astorga et al., 1994; Orozco et al., 1998).

Análisis comparativo de los genomas de los miembros del Clado SLCV con los de Curtovirus.

La notable similitud del dominio N-terminal de Rep de los virus examinados con dos miembros del género *Curtovirus*, Horseradish curly top virus (HrCTV) y Spinach severe curly top virus (SpiSCTV) (Figura 2.1) sugiere en primera instancia que los virus del linaje del SLCV se originaron de un evento de recombinación entre un curtovirus similar a los ya mencionados, y un begomovirus típico del NM. Sin embargo, el exámen exhaustivo de las secuencias de esos virus llevaron a las siguientes conclusiones:

- 1) En este escenario, la recombinación de los curtovirus habría abarcado solo los primeros 162 codones de *Rep*; en vez de ello, lo que se observa es una alta similitud a lo largo de todo el gen *Rep*. Por la misma razón, la proteína AC4* de esos curtovirus es muy similar a la proteína homóloga de los BGVs del linaje SLCV, y muy diferente a la proteína C4 de otros curtovirus (diversas cepas de Beet curly top virus, BCTV).
- 2) El análisis de la proteína C2 de HrCTV y SpiSCTV con la de begomovirus de varios continentes mostró una pocas semejanzas en su extremo N-terminal, sin tener una identidad global realmente significativa. En cambio, la proteína C2 de los anteriores curtovirus con las distintas cepas de BCTV (la tercera especie de curtovirus con múltiples cepas) no mostró semejanzas significativas a nivel de secuencia de aminoácidos.

Considerados en conjunto, los resultados de nuestro análisis de secuencia sistemático no proporcionó ninguna evidencia en apoyo a la suposición de que los miembros del Clado del SLCV adquirieron su módulo genómico distintivo (unos 670 pb) a partir de un curtovirus, o un geminivirus que se haya caracterizado al presente. Por el contrario, los datos apoyan fuertemente la hipótesis de que HrCTV y SpiSCTV adquirieron sus genes *Rep* y *AC4** atípicos como consecuencia de una recombinación con un begomovirus del grupo del SLCV.

Las anteriores reflexiones conducen a proponer un esquema de la evolución del gen *Rep* y C2 de begomovirus y curtovirus (**Figura 2.6**) que se centra en dos secuencias diferenciales altamente conservadas, que se encuentran en la vecindad del dominio de interacción con el regulador del ciclo celular “Retinoblastoma-related

protein” (RbRP) (Kong et al. 2000; Argüello-Astorga et al., 2007) y el dominio de oligomerización de Rep (Orozco et al., 2000)

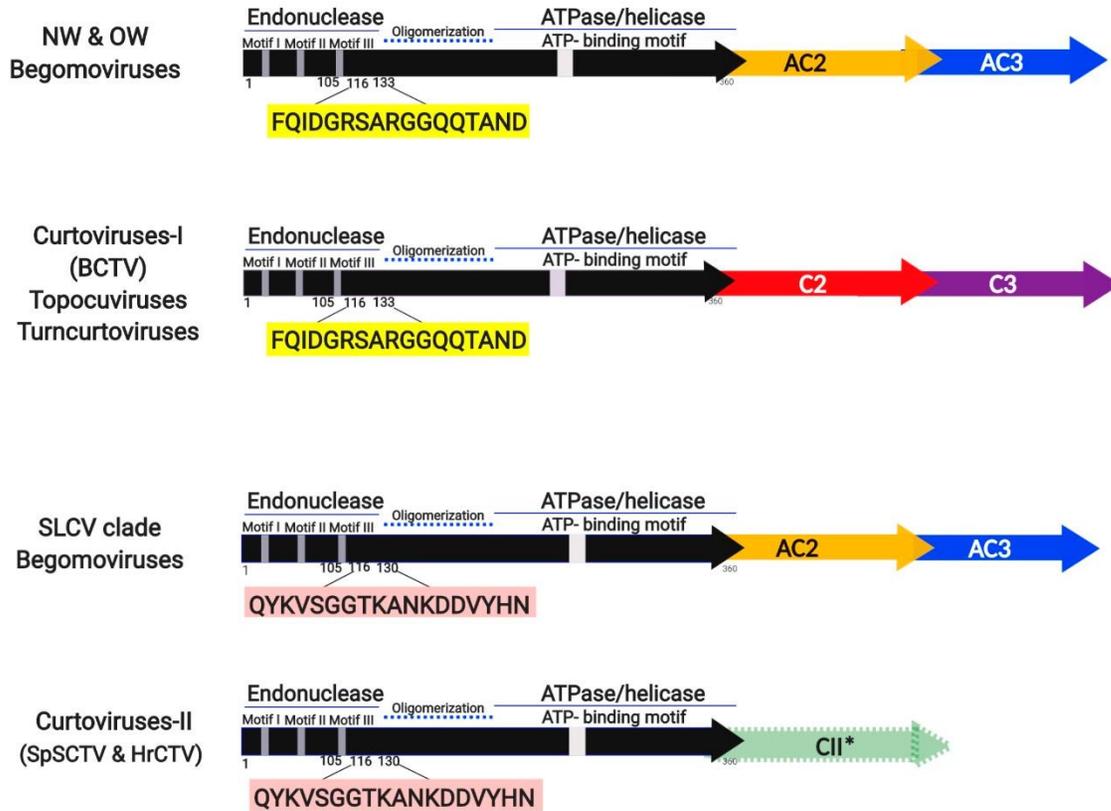


Figura 2.6- Representación simplificada de los genes (ORFs) que codifican a Rep y otras proteínas en el sentido complementario del genoma de los begomovirus y diversos géneros de geminivirus. Las secuencias de 17-18 residuos aminoácidos altamente conservadas (“firmas peptídicas”, resaltadas en amarillo o magenta) en el segmento 112-130, son características de los distintos subgrupos virales analizados en este estudio. Dichas firmas se encuentran en la vecindad del Motivo conservado RCR-III y cercano al dominio de oligomerización. Los genes C2 y C3 de los miembros de diferentes linajes presentan colores diferentes para enfatizar la divergencia estructural o funcional de sus homólogos en los begomovirus. La flecha verde mate de los curtovirus HrCTV y SpiSCTV representa el ORF C2 (o CII) altamente divergente del resto de las proteínas posicionalmente análogas de estos virus, cuya función se desconoce. Los begomovirus del Nuevo y el Viejo Mundo se denotan como NW y OW, respectivamente.

Conclusiones

Los miembros del linaje del SLCV (cerca de 40 especies distribuidas en todas las áreas cálidas de las Américas) codifican proteínas Rep y AC4* que presentan secuencias altamente divergentes de sus presuntos homólogos en otros begomovirus. Lo que logramos establecer en este análisis es que se trata de un subgrupo de virus que poseen un módulo de replicación de origen recombinante, de cerca de ~670nt, que codifica 160 residuos aminoácidos del extremo N-terminal de una Rep muy divergente, un origen de replicación atípico con 5 a 6 iterones con un consenso distinto a los iterones de otros geminivirus, y un putativo sitio de unión para un factor transcripcional (E2F) involucrado en la progresión del ciclo celular, el cual se encuentra íntimamente asociado al origen de replicación viral. Adicionalmente, estos virus codifican una proteína AC4* que no parece ser homóloga a las proteínas C4/AC4 de otros geminivirus. Existen dos curtovirus con elevada similitud a los virus del clado del SLCV, pero el análisis comparado de secuencias demostró que esos curtovirus recombinaron con un begomovirus del linaje del SLCV. Por tanto, queda abierta la cuestión de cual fue el virus donador del módulo replicativo atípico antes mencionado. Mediante el análisis bioinformático de bases de datos genómicos de virus y plantas, usando el dominio endonucleolítico de la Rep de esos BGVs, se lograron identificar EVS y virus no relacionados a geminivirus que codifican proteínas Rep con similitudes significativas a los virus en cuestión. De particular interés fue el hallazgo de un virus descubierto en un estudio metagenómico de una comunidad de invertebrados en el fondo de un Lago de Nueva Zelanda, que no sólo codifica una proteína Rep con similitud significativa en su dominio endonucleolítico a miembros del clado del SLCV, sino que además posee un origen de replicación con tres iterones en tándem que son idénticos a los iterones de los begomovirus mencionados. Estas observaciones sugieren que en la naturaleza existen virus de animales o de plantas que codifican proteínas Rep similares a los de los virus del linaje del SLCV, y que probablemente un begomovirus del NM adquirió, por recombinación, su módulo de replicación de algún virus de ese tipo, el cual progresivamente se adaptó a funcionar en el contexto de un genoma bipartita replicándose en el núcleo de la célula vegetal, y adquiriendo gradualmente nuevas funciones. En ausencia de evidencia más extensa y definida, esa es la mejor hipótesis que podemos proponer acerca del origen evolutivo de esos begomovirus atípicos, que hemos analizado exhaustivamente en este trabajo de tesis, en paralelo al tema central de la misma.

BIBLIOGRAFÍA

- Andrade EC, Manhani GG, Alfenas PF, Calegario RF, Fontes EPB, Zerbini FM. Tomato yellow spot virus, a tomato-infecting begomovirus from Brazil with a closer relationship to viruses from *Sida* sp., forms pseudorecombinants with begomoviruses from tomato but not from *Sida*. *J Gen Virol.* 2006;87(Pt 12):3687-3696. doi:10.1099/vir.0.82279-0
- Argüello-Astorga GR, Ruiz-Medrano R. An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol.* 2001;146(8):1465-1485. doi:10.1007/s007050170072
- Argüello-Astorga GR, Guevara-González RG, Herrera-Estrella LR, Rivera-Bustamante RF. Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. *Virology.* 1994;203(1):90-100. doi:10.1006/viro.1994.1458
- Argüello-Astorga G, Herrera-Estrella L, Rivera-Bustamante R. Experimental and theoretical definition of geminivirus origin of replication. *Plant Mol Biol.* 1994;26(2):553-556. doi:10.1007/BF00013742
- Arguello-Astorga G, Lopez-Ochoa L, Kong LJ, Orozco BM, Settlege SB, Hanley-Bowdoin L. A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein. *J Virol.* 2004;78(9):4817-4826. doi:10.1128/jvi.78.9.4817-4826.2004
- Avalos-Calleros JA, Pastor-Palacios G, Bolaños-Martínez OC, et al. Two strains of a novel begomovirus encoding Rep proteins with identical β 1 strands but different β 5 strands are not compatible in replication. *Arch Virol.* 2021;166(6):1691-1709. doi:10.1007/s00705-021-05066-2
- Bañuelos Hernández, Bernardo (2012) Identificación de los determinantes de especificidad de la proteína iniciadora de la replicación de geminivirus por mutagénesis dirigida. Junio 2012. Tesis Doctoral. IPICYT.
- Bisaro DM. Silencing suppression by geminivirus proteins. *Virology.* 2006;344(1):158-168. doi:10.1016/j.virol.2005.09.041
- Lee J, Shin MK, Ryu DK, Kim S, Ryu WS. Insertion and deletion mutagenesis by overlap extension PCR. *Methods Mol Biol.* 2010;634:137-146. doi:10.1007/978-1-60761-652-8_10
- Bridson RW, Markham PG. Complementation of bipartite begomovirus movement functions by topocoviruses and curtoviruses. *Arch Virol.* 2001;146(9):1811-1819. doi:10.1007/s007050170067
- Brown JK, Zerbini FM, Navas-Castillo J, et al. Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Arch Virol.* 2015;160(6):1593-1619. doi:10.1007/s00705-015-2398-y
- Buchmann RC, Asad S, Wolf JN, Mohannath G, Bisaro DM. Geminivirus AL2 and L2 proteins suppress transcriptional gene silencing and cause genome-wide reductions in cytosine methylation. *J Virol.* 2009;83(10):5005-5013. doi:10.1128/JVI.01771-08
- Campos-Olivas R, Louis JM, Clerot D, Gronenborn B, Gronenborn AM. The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. *Proc Natl Acad Sci U S A.* 2002;99(16):10310-10315. doi:10.1073/pnas.152342699

- Castellano MM, Sanz-Burgos AP, Gutiérrez C. Initiation of DNA replication in a eukaryotic rolling-circle replicon: identification of multiple DNA-protein complexes at the geminivirus origin. *J Mol Biol.* 1999;290(3):639-652. doi:10.1006/jmbi.1999.2916
- Chabouté ME, Clément B, Sekine M, Philipps G, Chaubet-Gigot N. Cell cycle regulation of the tobacco ribonucleotide reductase small subunit gene is mediated by E2F-like elements. *Plant Cell.* 2000 Oct;12(10):1987-2000.
- Chakraborty S, Vanitharani R, Chattopadhyay B, Fauquet CM. Supervirulent pseudorecombination and asymmetric synergism between genomic components of two distinct species of begomovirus associated with severe tomato leaf curl disease in India. *J Gen Virol.* 2008;89(Pt 3):818-828. doi:10.1099/vir.0.82873-0
- Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B. Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol.* 2013;11(8):525-538. doi:10.1038/nrmicro3067
- Chatterji A, Padidam M, Beachy RN, Fauquet CM. Identification of replication specificity determinants in two strains of tomato leaf curl virus from New Delhi. *J Virol.* 1999;73(7):5481-5489. doi:10.1128/JVI.73.7.5481-5489.1999
- Choi IR, Stenger DC. Strain-specific determinants of beet curly top geminivirus DNA replication. *Virology.* 1995;206(2):904-912. doi:10.1006/viro.1995.1013
- Choi IR, Stenger DC. The strain-specific cis-acting element of beet curly top geminivirus DNA replication maps to the directly repeated motif of the ori. *Virology.* 1996;226(1):122-126. doi:10.1006/viro.1996.0634
- Choudhury NR, Malik PS, Singh DK, Islam MN, Kaliappan K, Mukherjee SK. The oligomeric Rep protein of Mungbean yellow mosaic India virus (MYMIV) is a likely replicative helicase. *Nucleic Acids Res.* 2006;34(21):6362-6377. doi:10.1093/nar/gkl903
- Clérot D, Bernardi F. DNA helicase activity is associated with the replication initiator protein rep of tomato yellow leaf curl geminivirus. *J Virol.* 2006;80(22):11322-11330. doi:10.1128/JVI.00924-06
- Czosnek H, Hariton-Shalev A, Sobol I, Gorovits R, Ghanim M. The Incredible Journey of Begomoviruses in Their Whitefly Vector. *Viruses.* 2017;9(10):273. doi:10.3390/v9100273
- Dayaram A, Galatowitsch ML, Argüello-Astorga GR, van Bysterveldt K, Kraberger S, Stainton D, Harding JS, Roumagnac P, Martin DP, Lefeuvre P, Varsani A. Diverse circular replication-associated protein encoding viruses circulating in invertebrates within a lake ecosystem. *Infect Genet Evol.* 2016 Apr;39:304-316. doi: 10.1016/j.meegid.2016.02.011.
- De Barro PJ, Liu SS, Boykin LM, Dinsdale AB. Bemisia tabaci: a statement of species status. *Annu Rev Entomol.* 2011;56:1-19. doi:10.1146/annurev-ento-112408-085504
- Dellaporta, S.L.; Wood, J.; Hicks, J.B. A plant DNA miniprep: Version II. *Plant Mol. Biol. Report.* 1983;19–21.
- Desbiez C, David C, Mettouchi A, Laufs J, Gronenborn B. Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication [published correction appears in *Proc Natl Acad Sci U S A* 1995 Nov 21;92(24):11322]. *Proc Natl Acad Sci U S A.* 1995;92(12):5640-5644. doi:10.1073/pnas.92.12.5640
- Donson J, Morris-Krsinich BA, Mullineaux PM, Boulton MI, Davies JW. A putative primer for second-strand DNA synthesis of maize streak virus is virion-associated. *EMBO J.* 1984;3(13):3069-3073. doi:10.1002/j.1460-2075.1984.tb02260.x

- Eagle PA, Orozco BM, Hanley-Bowdoin L. A DNA sequence required for geminivirus replication also mediates transcriptional regulation. *Plant Cell*. 1994;6(8):1157-1170. doi:10.1105/tpc.6.8.1157
- Elmer JS, Brand L, Sunter G, Gardiner WE, Bisaro DM, Rogers SG. Genetic analysis of the tomato golden mosaic virus. II. The product of the AL1 coding sequence is required for replication. *Nucleic Acids Res*. 1988;16(14B):7043-7060. doi:10.1093/nar/16.14.7043
- Elmer JS, Sunter G, Gardiner WE, et al. Agrobacterium-mediated inoculation of plants with tomato golden mosaic virus DNAs. *Plant Mol Biol*. 1988;10(3):225-234. doi:10.1007/BF00027399
- Fiallo-Olivé E, Lett JM, Martin DP, et al. ICTV Virus Taxonomy Profile: *Geminiviridae* 2021. *J Gen Virol*. 2021;102(12):001696. doi:10.1099/jgv.0.001696
- Fondong VN. Geminivirus protein structure and function. *Mol Plant Pathol*. 2013;14(6):635-649. doi:10.1111/mpp.12032
- Fondong VN. The Ever-Expanding Role of C4/AC4 in Geminivirus Infection: Punching above Its Weight? *Mol Plant*. 2019 Feb 4;12(2):145-147. doi: 10.1016/j.molp.2018.12.006.
- Fontes EP, Eagle PA, Sipe PS, Luckow VA, Hanley-Bowdoin L. Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *J Biol Chem*. 1994;269(11):8459-8465.
- Fontes EP, Gladfelter HJ, Schaffer RL, Petty IT, Hanley-Bowdoin L. Geminivirus replication origins have a modular organization. *Plant Cell*. 1994;6(3):405-416. doi:10.1105/tpc.6.3.405
- Fontes EP, Luckow VA, Hanley-Bowdoin L. A geminivirus replication protein is a sequence-specific DNA binding protein. *Plant Cell*. 1992;4(5):597-608. doi:10.1105/tpc.4.5.597
- García-Andrés S, Tomás DM, Sánchez-Campos S, Navas-Castillo J, Moriones E. Frequent occurrence of recombinants in mixed infections of tomato yellow leaf curl disease-associated begomoviruses. *Virology*. 2007;365(1):210-219. doi:10.1016/j.virol.2007.03.045
- Gregorio-Jorge J, Bernal-Alcocer A, Bañuelos-Hernández B, et al. Analysis of a new strain of Euphorbia mosaic virus with distinct replication specificity unveils a lineage of begomoviruses with short Rep sequences in the DNA-B intergenic region. *Virology*. 2010;7:275. doi:10.1186/1743-422X-7-275
- Gutierrez C. Geminivirus DNA replication. *Cell Mol Life Sci*. 1999;56(3-4):313-329. doi:10.1007/s000180050433
- Hanley-Bowdoin L, Bejarano ER, Robertson D, Mansoor S. Geminiviruses: masters at redirecting and reprogramming plant processes. *Nat Rev Microbiol*. 2013;11(11):777-788. doi:10.1038/nrmicro3117
- Hesketh EL, Saunders K, Fisher C, et al. The 3.3 Å structure of a plant geminivirus using cryo-EM. *Nat Commun*. 2018;9(1):2369. doi:10.1038/s41467-018-04793-6
- Higo K, Ugawa Y, Iwamoto M, Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res*. 1999 Jan 1;27(1):297-300. doi: 10.1093/nar/27.1.297.
- Hipp K, Grimm C, Jeske H, Böttcher B. Near-Atomic Resolution Structure of a Plant Geminivirus Determined by Electron Cryomicroscopy. *Structure*. 2017;25(8):1303-1309.e3. doi:10.1016/j.str.2017.06.013

- Hung HC, Petty ITD. Functional equivalence of late gene promoters in bean golden mosaic virus with those in tomato golden mosaic virus. *J Gen Virol.* 2001;82(Pt 3):667-672. doi:10.1099/0022-1317-82-3-667
- Hull, R. Chapter 2 - Plant Viruses and Their Classification, Editor(s): Roger Hull, *Plant Virology (Fifth Edition)*, Academic Press,2014;15-68, doi.org/10.1016/B978-0-12-384871-0.00002-9.
- Ilyina TV, Koonin EV. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria. *Nucleic Acids Res.* 1992;20(13):3279-3285. doi:10.1093/nar/20.13.3279
- Ingham DJ, Pascal E, Lazarowitz SG. Both bipartite geminivirus movement proteins define viral host range, but only BL1 determines viral pathogenicity. *Virology.* 1995;207(1):191-204. doi:10.1006/viro.1995.1066
- Jeske H. Geminiviruses. *Curr Top Microbiol Immunol.* 2009;331:185-226. doi:10.1007/978-3-540-70972-5_11
- Jeske H, Lütgemeier M, Preiss W. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO J.* 2001;20(21):6158-6167. doi:10.1093/emboj/20.21.6158
- Jupin I, Hericourt F, Benz B, Gronenborn B. DNA replication specificity of TYLCV geminivirus is mediated by the amino-terminal 116 amino acids of the Rep protein. *FEBS Lett.* 1995;362(2):116-120. doi:10.1016/0014-5793(95)00221-t
- Kim KI, Sunter G, Bisaro DM, Chung IS. Improved expression of recombinant GFP using a replicating vector based on Beet curly top virus in leaf-disks and infiltrated *Nicotiana benthamiana* leaves. *Plant Mol Biol.* 2007;64(1-2):103-112. doi:10.1007/s11103-007-9137-z
- Kong LJ, Orozco BM, Roe JL, et al. A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *EMBO J.* 2000;19(13):3485-3495. doi:10.1093/emboj/19.13.3485
- Krupovic M, Varsani A, Kazlauskas D, et al. *Cressdnaviricota*: a Virus Phylum Unifying Seven Families of Rep-Encoding Viruses with Single-Stranded, Circular DNA Genomes. *J Virol.* 2020;94(12):e00582-20. doi:10.1128/JVI.00582-20
- Kumar RV, Prasanna HC, Singh AK, Ragunathan D, Garg GK, Chakraborty S. Molecular genetic analysis and evolution of begomoviruses and betasatellites causing yellow mosaic disease of bhendi. *Virus Genes.* 2017;53(2):275-285. doi:10.1007/s11262-016-1414-y
- Kushwaha NK, Mansi, Chakraborty S. The replication initiator protein of a geminivirus interacts with host monoubiquitination machinery and stimulates transcription of the viral genome [published correction appears in PLoS Pathog. 2018 Aug 29;14(8):e1007281]. *PLoS Pathog.* 2017;13(8):e1006587. Published 2017 Aug 31. doi:10.1371/journal.ppat.1006587
- Laufs J, Traut W, Heyraud F, et al. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc Natl Acad Sci U S A.* 1995;92(9):3879-3883. doi:10.1073/pnas.92.9.3879
- Lazarowitz SG, Wu LC, Rogers SG, Elmer JS. Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. *Plant Cell.* 1992;4(7):799-809. doi:10.1105/tpc.4.7.799
- Li KB. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics.* 2003;19(12):1585-1586. doi:10.1093/bioinformatics/btg192

- Londoño A, Riego-Ruiz L, Argüello-Astorga GR. DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. *Arch Virol*. 2010;155(7):1033-1046. doi:10.1007/s00705-010-0674-4
- Mauricio-Castillo JA, Torres-Herrera SI, Cárdenas-Conejo Y, Pastor-Palacios G, Méndez-Lozano J, Argüello-Astorga GR. A novel begomovirus isolated from sida contains putative cis- and trans-acting replication specificity determinants that have evolved independently in several geographical lineages. *Arch Virol*. 2014;159(9):2283-2294. doi:10.1007/s00705-014-2073-8v
- Medina-Puche L, Orílio AF, Zerbini FM, Lozano-Durán R. Small but mighty: Functional landscape of the versatile geminivirus-encoded C4 protein. *PLoS Pathog*. 2021;17(10):e1009915. doi:10.1371/journal.ppat.1009915
- Melgarejo TA, Kon T, Rojas MR, Paz-Carrasco L, Zerbini FM, Gilbertson RL. Characterization of a new world monopartite begomovirus causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in geminivirus evolution. *J Virol*. 2013;87(10):5397-5413. doi:10.1128/JVI.00234-13
- Mengshi Wu, Eduardo R. Bejarano, Araceli G. Castillo, Rosa Lozano-Durán, Chapter 20 - Geminivirus DNA replication in plants, Editor(s): R.K. Gaur, Pradeep Sharma, Henryk Czosnek, Geminivirus : Detection, Diagnosis and Management, *Academic Press*, 2022, Pages 323-346, doi.org/10.1016/B978-0-323-90587-9.00038-9.
- Nash TE, Dallas MB, Reyes MI, Buhman GK, Ascencio-Ibañez JT, Hanley-Bowdoin L. Functional analysis of a novel motif conserved across geminivirus Rep proteins. *J Virol*. 2011;85(3):1182-1192. doi:10.1128/JVI.02143-10
- Navas-Castillo J, Fiallo-Olivé E, Sánchez-Campos S. Emerging virus diseases transmitted by whiteflies. *Annu Rev Phytopathol*. 2011;49:219-248. doi:10.1146/annurev-phyto-072910-095235
- Orozco BM, Gladfelter HJ, Settlege SB, Eagle PA, Gentry RN, Hanley-Bowdoin L. Multiple cis elements contribute to geminivirus origin function. *Virology*. 1998 15;242(2):346-56. doi:10.1006/viro.1997.9013.
- Orozco BM, Hanley-Bowdoin L. Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *J Biol Chem*. 1998;273(38):24448-24456. doi:10.1074/jbc.273.38.24448
- Orozco BM, Kong LJ, Batts LA, Elledge S, Hanley-Bowdoin L. The multifunctional character of a geminivirus replication protein is reflected by its complex oligomerization properties. *J Biol Chem*. 2000 Mar 3;275(9):6114-22. doi:10.1074/jbc.275.9.6114.
- Paprotka T, Deuschle K, Metzler V, Jeske H. Conformation-selective methylation of geminivirus DNA. *J Virol*. 2011;85(22):12001-12012. doi:10.1128/JVI.05567-11
- Peyret H, Lomonosoff GP. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol Biol*. 2013;83(1-2):51-58. doi:10.1007/s11103-013-0036-1
- Pilartz M, Jeske H. Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. *Virology*. 1992;189(2):800-802. doi:10.1016/0042-6822(92)90610-2
- Porta C, Lomonosoff GP. Use of viral replicons for the expression of genes in plants. *Mol Biotechnol*. 1996;5(3):209-221. doi:10.1007/BF02900359
- Preiss W, Jeske H. Multitasking in replication is common among geminiviruses. *J Virol*. 2003;77(5):2972-2980. doi:10.1128/jvi.77.5.2972-2980.2003

- Raja P, Wolf JN, Bisaro DM. RNA silencing directed against geminiviruses: post-transcriptional and epigenetic components. *Biochim Biophys Acta*. 2010;1799(3-4):337-351. doi:10.1016/j.bbagr.2010.01.004
- Ramos PL, Guevara-González RG, Peral R, et al. Tomato mottle Taino virus pseudorecombines with PYMV but not with ToMoV: implications for the delimitation of cis- and trans-acting replication specificity determinants. *Arch Virol*. 2003;148(9):1697-1712. doi:10.1007/s00705-003-0136-3
- Revington GN, Sunter G, Bisaro DM. DNA sequences essential for replication of the B genome component of tomato golden mosaic virus. *Plant Cell*. 1989;1(10):985-992. doi:10.1105/tpc.1.10.985
- Rizvi I, Choudhury NR, Tuteja N. Insights into the functional characteristics of geminivirus rolling-circle replication initiator protein and its interaction with host factors affecting viral DNA replication. *Arch Virol*. 2015;160(2):375-387. doi:10.1007/s00705-014-2297-7
- Rodríguez-Gandarilla MG, Rodríguez-Negrete EA, Rivera-Bustamante RF. Superinfection by PHYVV Alters the Recovery Process in PepGMV-Infected Pepper Plants. *Viruses*. 2020;12(3):286. Published 2020 Mar 5. doi:10.3390/v12030286
- Rojas MR, Hagen C, Lucas WJ, Gilbertson RL. Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. *Annu Rev Phytopathol*. 2005;43:361-394. doi:10.1146/annurev.phyto.43.040204.135939
- Rojas MR, Macedo MA, Maliano MR, et al. World Management of Geminiviruses. *Annu Rev Phytopathol*. 2018;56:637-677. doi:10.1146/annurev-phyto-080615-100327
- Romay G, Chirinos DT, Geraud-Pouey F, et al. Molecular and biological characterization of a new Tomato mild yellow leaf curl Aragua virus strain producing severe symptoms in tomato. *Virus Genes*. 2017;53(6):939-942. doi:10.1007/s11262-017-1478-3
- Ruhel R, Chakraborty S. Multifunctional roles of geminivirus encoded replication initiator protein. *Virusdisease*. 2019;30(1):66-73. doi:10.1007/s13337-018-0458-0
- Rybicki EP. A phylogenetic and evolutionary justification for three genera of Geminiviridae. *Arch Virol*. 1994;139(1-2):49-77. doi: 10.1007/BF01309454.
- Sanchez-Chavez S, Regla-Marquez CF, Cardenas-Conejo ZE, Garcia-Rodriguez DA, Centeno-Leija S, Serrano-Posada H, Liñan-Rico A, Partida-Palacios BL, Cardenas-Conejo Y. First report of begomoviruses infecting *Cucumis sativus* L. in North America and identification of a proposed new begomovirus species. *PeerJ*. 2020 Jul 10;8:e9245. doi: 10.7717/peerj.9245.
- Sanderfoot AA, Ingham DJ, Lazarowitz SG. A viral movement protein as a nuclear shuttle. The geminivirus BR1 movement protein contains domains essential for interaction with BL1 and nuclear localization. *Plant Physiol*. 1996;110(1):23-33. doi:10.1104/pp.110.1.23
- Saunders K, Briddon RW, Stanley J. Replication promiscuity of DNA-beta satellites associated with monopartite begomoviruses; deletion mutagenesis of the Ageratum yellow vein virus DNA-beta satellite localizes sequences involved in replication. *J Gen Virol*. 2008;89(Pt 12):3165-3172. doi:10.1099/vir.0.2008/003848-0
- Saunders K, Lucy A, Stanley J. DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Res*. 1991;19(9):2325-2330. doi:10.1093/nar/19.9.2325

- Saunders K, Salim N, Mali VR, et al. Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. *Virology*. 2002;293(1):63-74. doi:10.1006/viro.2001.1251
- Settlage SB, See RG, Hanley-Bowdoin L. Geminivirus C3 protein: replication enhancement and protein interactions. *J Virol*. 2005;79(15):9885-9895. doi:10.1128/JVI.79.15.9885-9895.2005
- Sharma P, Ikegami M. Tomato leaf curl Java virus V2 protein is a determinant of virulence, hypersensitive response and suppression of posttranscriptional gene silencing. *Virology*. 2010;396(1):85-93. doi:10.1016/j.virol.2009.10.012
- Shimada-Beltrán H, Rivera-Bustamante RF. Early and late gene expression in pepper huasteco yellow vein virus. *J Gen Virol*. 2007;88(Pt 11):3145-3153. doi:10.1099/vir.0.83003-0
- Silva JCF, Carvalho TFM, Basso MF, et al. Geminivirus data warehouse: a database enriched with machine learning approaches. *BMC Bioinformatics*. 2017;18(1):240. doi:10.1186/s12859-017-1646-4
- Singh DK, Islam MN, Choudhury NR, Karjee S, Mukherjee SK. The 32 kDa subunit of replication protein A (RPA) participates in the DNA replication of Mung bean yellow mosaic India virus (MYMIV) by interacting with the viral Rep protein. *Nucleic Acids Res*. 2007;35(3):755-770. doi:10.1093/nar/gkl1088
- Singh DK, Malik PS, Choudhury NR, Mukherjee SK. MYMIV replication initiator protein (Rep): roles at the initiation and elongation steps of MYMIV DNA replication. *Virology*. 2008;380(1):75-83. doi:10.1016/j.virol.2008.07.010
- Teng K, Chen H, Lai J, et al. Involvement of C4 protein of beet severe curly top virus (family Geminiviridae) in virus movement. *PLoS One*. 2010;5(6):e11280. doi:10.1371/journal.pone.0011280
- Torres-Herrera SI. Caracterización de nuevos begomovirus de un linaje que evolucionó por recombinación con un geminivirus de origen desconocido. Marzo 2015. Tesis Doctoral, IPICYT.
- Torres-Herrera SI, Romero-Osorio A, Moreno-Valenzuela O, et al. A Lineage of Begomoviruses Encode Rep and AC4 Proteins of Enigmatic Ancestry: Hints on the Evolution of Geminiviruses in the New World. *Viruses*. 2019;11(7):644. doi:10.3390/v11070644
- Vanitharani R, Chellappan P, Fauquet CM. Geminiviruses and RNA silencing. *Trends Plant Sci*. 2005;10(3):144-151. doi:10.1016/j.tplants.2005.01.005
- Varsani A, Navas-Castillo J, Moriones E, et al. Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch Virol*. 2014;159(8):2193-2203. doi:10.1007/s00705-014-2050-2
- Vega-Rocha S, Gronenborn B, Gronenborn AM, Campos-Olivas R. Solution structure of the endonuclease domain from the master replication initiator protein of the nanovirus faba bean necrotic yellows virus and comparison with the corresponding geminivirus and circovirus structures. *Biochemistry*. 2007;46(21):6201-6212. doi:10.1021/bi700159q
- Venkataravanappa V, Lakshminarayana Reddy CN, Jalali S, Krishna Reddy M. Association of tomato leaf curl New Delhi virus DNA-B with bhendi yellow vein mosaic virus in okra showing yellow vein mosaic disease symptoms. *Acta Virol*. 2015;59(2):125-139. doi:10.4149/av_2015_02_125

Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GT, Gruissem W, Van de Peer Y, Inzé D, De Veylder L. Genome-wide identification of potential plant E2F target genes. *Plant Physiol.* 2005 Sep;139(1):316-28. doi: 10.1104/pp.105.066290. Epub 2005 Aug 26.

Xu X, Qian Y, Wang Y, Li Z, Zhou X. Iterons Homologous to Helper Geminiviruses Are Essential for Efficient Replication of Betasatellites. *J Virol.* 2019;93(5):e01532-18. Published 2019 Feb 19. doi:10.1128/JVI.01532-18

Zerbini FM, Briddon RW, Idris A, et al. ICTV Virus Taxonomy Profile: Geminiviridae. *J Gen Virol.* 2017;98(2):131-133. doi:10.1099/jgv.0.000738

Zeng R, Liu X, Yang C, Lai J. Geminivirus C4: Interplaying with Receptor-like Kinases. *Trends Plant Sci.* 2018;23(12):1044-1046. doi:10.1016/j.tplants.2018.09.003

Zhang W, Olson NH, Baker TS, et al. Structure of the Maize streak virus geminate particle. *Virology.* 2001;279(2):471-477. doi:10.1006/viro.2000.0739

Zhou Y, Rojas MR, Park MR, Seo YS, Lucas WJ, Gilbertson RL. Histone H3 interacts and colocalizes with the nuclear shuttle protein and the movement protein of a geminivirus. *J Virol.* 2011;85(22):11821-11832. doi:10.1128/JVI.00082-11

Anexo 1

1. Artículo publicado (primer autoría compartida*):

Torres-Herrera, S.I.*; **Romero-Osorio, A***; Moreno-Valenzuela, O.; Pastor-Palacios, G.; Cardenas-Conejo, Y.; Ramírez-Prado, J.H.; Riego-Ruiz, L.; Minero-García, Y.; Ambriz-Granados, S.; Argüello-Astorga, G.R. "A Lineage of Begomoviruses Encode Rep and AC4 Proteins of Enigmatic Ancestry: Hints on the Evolution of Geminiviruses in the New World." ***Viruses***. 2019; 11 (7): E644 doi: [10.3390/v11070644](https://doi.org/10.3390/v11070644) (Open Access)

Anexo 2

Material Suplementario

Tabla S1. Oligonucleótidos utilizados para generar el sitio NgoMIV en *Rep*

NOMBRE	SECUENCIA 5'→3'	LONGITUD	USO
FNgoMIV-3	CTT CTT ATT ACA AAA ATG CCG GCA CCA AAG AAA TTT AGA GTT	43	generar sitio de corte NgoMIV en gen <i>rep</i>
RNgoMIV-3	GAA CTC TAA ATT CTC TTG GTG CCG GCA TTT TTG TAA TAA GAA G	43	generar sitio de corte NgoMIV en gen <i>rep</i>
M13For	GTTTTCCAGTCACGTTGTA	20	Universal-donado por LANBAMA
M13 ReV	TTGTGAGCGGATAACAATTC	21	Universal-donado por LANBAMA

Tabla S2. Casetes sintéticos con los codones 5, 8 y 10 del gen *Rep* de ToMoV

Nombre	Secuencia 5'→3'	Tamaño (nt)	característica
FC1-KRN	CCG GCA CCA <u>AAG</u> AAA TTT <u>AGA</u> GTT <u>ACC</u> TCA AAG AAC T	37	casete para cambios en aa 5, 8 y 10 (KRN)
RC1-KRN	AGT TCT TTG AGT TAA CTC TAA ATT TCT TTG GTG	33	casete para cambios en aa 5, 8 y 10 (KRN)
FC2-KKS	CCG GCA CCA <u>AAG</u> AAA TTT <u>AAA</u> GTT <u>TCA</u> TCA AAG AAC T	37	casete para cambios en aa 5, 8 y 10(KKS)
RC2-KKS	AGT TCT TTG ATG AAA CTT TAA ATT TCT TTG GTG	33	casete para cambios en aa 5, 8 y 10(KKS)
FC3-KKQ	CCG GCA CCA <u>AAG</u> AAA TTT <u>AAA</u> GTT <u>CAG</u> TCA AAG AAC T	37	casete para cambios en aa 5, 8 y 10(KKQ)
RC3-KKQ	AGT TCT TTG ACT GAA CTT TAA ATT TCT TTG GTG	33	casete para cambios en aa 5, 8 y 10(KKQ)
FC4-GSK	CCG GCA CCA <u>GGA</u> AAA TTT <u>TCA</u> GTT <u>AAG</u> TCA AAG AAC T	37	casete para cambios en aa 5, 8 y 10 (GSK)
RC4-GSK	AGT TCT TTG ACT TAA CTG AAA ATT TTC CTG GTG	33	casete para cambios en aa 5, 8 y 10 (GSK)
FC5-KLN	CCG GCA CCA <u>AAG</u> AAA TTT <u>CTT</u> GTT <u>AAC</u> TCA AAG AAC T	37	casete para cambios en aa 5, 8 y 10 (KLN)
RC5-KLN	AGT TCT TTG AGT TAA CAA GAA ATT TCT TTG GTG	33	casete para cambios en aa 5, 8 y 11(KLN)

FC6-KQN	CCG GCA CCA <u>AAG</u> AAA TTT <u>CAG</u> GTT <u>AAC</u> TCA AAG AAC T	37	casete para cambios en aa 5, 8 y 12(KQN)
RC6-KQN	AGT TCT TTG AGT TAA CCT GAA ATT TCT TTC GTG	33	casete para cambios en aa 5, 8 y 13 (KQN)
FC8-NQK	CCG GCA CCA <u>AAC</u> AAA TTT <u>CAG</u> GTT <u>AAA</u> TCA AAG AAC T	37	casete para cambios en aa 5, 8 y 14 (NQK)
RC8-NQK	AGT TCT TTG ATT TAA CCT GAA ATT TGT TTG GTG	33	casete para cambios en aa 5, 8 y 15 (NQK)

Tabla S3. Oligonucleótidos para la mutación sitio dirigida de los residuos aa 69 y 71 de la proteína Rep de ToMoV.

Nombre	Secuencia 5' - →3'		Características
RuniPcil	TCG AAC TGA ACA AGC ACA TGA AGA TGA G	28	Oligo Reverso usado con los oligos siguientes
FC3V69	AGG TAA GTA <u>CGT</u> GTG CAC GAA TAA CAG ATT C	31	genera mutaciones en aa 69 de la clona KKQ
FC4N69	AGG TAA GTA <u>CAA</u> TTG CAC GAA TAA CAG ATT C	31	genera mutaciones en aa. 69 de la clona GSK
FC5KQ	AGG TAA GTA <u>CAA</u> ATG CCA GAA TAA CAG ATT	31	genera mutaciones en aa 69 Y 71 de la clona KLN y KQN
FC7CQ	AGG TAA GTA <u>CTG</u> CTG CCA GAA TAA CAG ATT C	31	genera mutaciones en aa 69 Y 71 de la clona KQN
FC8R71	AGG TAA GTA <u>CCA</u> ATG CAG AAA TAA CAG ATT C	31	genera mutaciones en aa. 71 de la clona NQK

Tabla S4. Mega-oligonucleótidos con iterones difentes a ToMoV

Fragmentos sintéticos con iterones	Secuencia	Longitud	Geminivirus que presentan el iterón
F-KKSQT-L	5'- CGCCACGTGTAGGAACCAGATTAATAGT TCTATAGAGGGATCTTATATATAAGTTAC CCAATACCCAATTGAATGAGCAAGTT TGAGAAGGC - 3'	94	CdTV
R-KKSQT-L	5'- CTTGCCGGCATT TTTGTAATAAGAAGGG GTACCCAGATGAGTTACCCAATTGAG CCTTCTCAAACCTTGCTCATTCAATTGGG GTATTGGG-3'	92	CdTV

F-NQKQR-L	5'- CGCCACGTGTAGGAACCAGATTAATAGT TCTATAGAGGGATCTTATATATAAGT GC ACCCAAGCACCCAATTGAATGAGCAAGT <u>TTGAGAAGG-3'</u>	94	BCTV
R-NQKQR-L	5'- CTTGCCGGCATT TTTGTAATAAGAAGGG GGCACCCAGATGAGTGCACCCAATTGA <u>GCCTTCTCAAACCTTGCTCATTCAATTGG</u> <u>GTGCTTGGGT-3'</u>	92	BCTV
F-KKQVT-L	5'- CGCCACGTGTAGGAACCAGATTAATAGT TCTATAGAGGGATCTTATATATAAGT GTC TCCAAGTCTCCAATTGAATGAGCAAGTT <u>TGAGAAGGC-3'</u>	94	ACMV
R-KKQVT-L	3'- CTTGCCGGCATT TTTGTAATAAGAAGGG GGTCTCCAGATGAGTGTCTCCAATTGAG <u>CCTTCTCAAACCTTGCTCATTCAATTGGA</u> <u>GACTTGGAG-3'</u>	93	ACMV
F-KRNQT-L	5'- CGCCACGTGTAGGAACCAGATTAATAGT TCTATAGAGGGATCTTATATATAAGT TAC ACCAATACACCAATTGAATGAGCAAGTT <u>TGAGAAGGC-3'</u>	94	PHYVV
R-KRNQT-L	5'- CTTGCCGGCATT TTTGTAATAAGAAGGG GTACACCAGATGAGTTACACCAATTGAG <u>CCTTCTCAAACCTTGCTCATTCAATTGGTG</u> <u>TATTGGT-3'</u>	92	PHYVV
F-KLNKQ-L	5'- CGCCACGTGTAGGAACCAGATTAATAGT TCTATAGAGGGATCTTATATATAAGT AG ACCAAAGACCCAATTGAATGAGCAAGT <u>TTGAGAAGGC-3'</u>	94	TYLChV
R-KLNKQ-L	5'- CTTGCCGGCATT TTTGTAATAAGAAGGG GAGACCCAGATGAGTAGACCCAATTGA GCCTTCTCAAACCTTGCTCATTCAATTGG GTCTTTGGG-3'	92	TYLChV
F-KQNCQ-L	5'- CGCCACGTGTAGGAACCAGATTAATAGT	94	TGMV

	TCTATAGAGGGATCTTATATATAAGTTCT ACCAATCTACCAATTGAATGAGCAAGTT <u>TGAGAAGGC-3'</u>		
R-KQNCQ-L	5'- CTTGCCGGCATT TTTTGT AATAAGAAGGG GACTACC AGATGAGT ACTACCA ATTGAG <u>CCTTCTCAA</u> ACTTGCTCATTCAATTGGTA <u>GATTGGT-3'</u>	92	TGMV
F-SiMSinV-L	5'- CCACGTGTAGGAACCAGATTAATAGTTC TATAGAGGGATCTTATATATAAGTT CCCC CAATCCCCCA ATTGAATGAGCAAGTTTG <u>AGAAGGC-3</u>	94	SiMSinV
R-SiMSinV-L	5'- <u>TGCCGGCATT TTTTGT AATAAGAAGGGGT</u> CCCCAGATGAGTTCCCCA ATTGAGC CTTCTCAA ACTTGCTCATTCAATTGGGG <u>GATTGGG-3'</u>	92	SiMSinV
F-OKYMMV-L	5'- CCACGTGTAGGAACCAGATTAATAGTTC TATAGAGGGATCTTATATATAAT GTGTAC CTGTGTACCA ATTGAATGAGCAAGTTTT <u>GAGAAGGCT-3'</u>	94	OKYMMV
R-OKYMMV-L	5'- <u>TGCCGGCATT TTTTGT AATAAGAAGGTGT</u> GTACCAGATGATGTGTACCA ATTGAGCC TTCTCAA ACTTGCTCATTCAATTGGTAC <u>ACAGGTA-3'</u>	92	OKYMMV

ANEXO 3

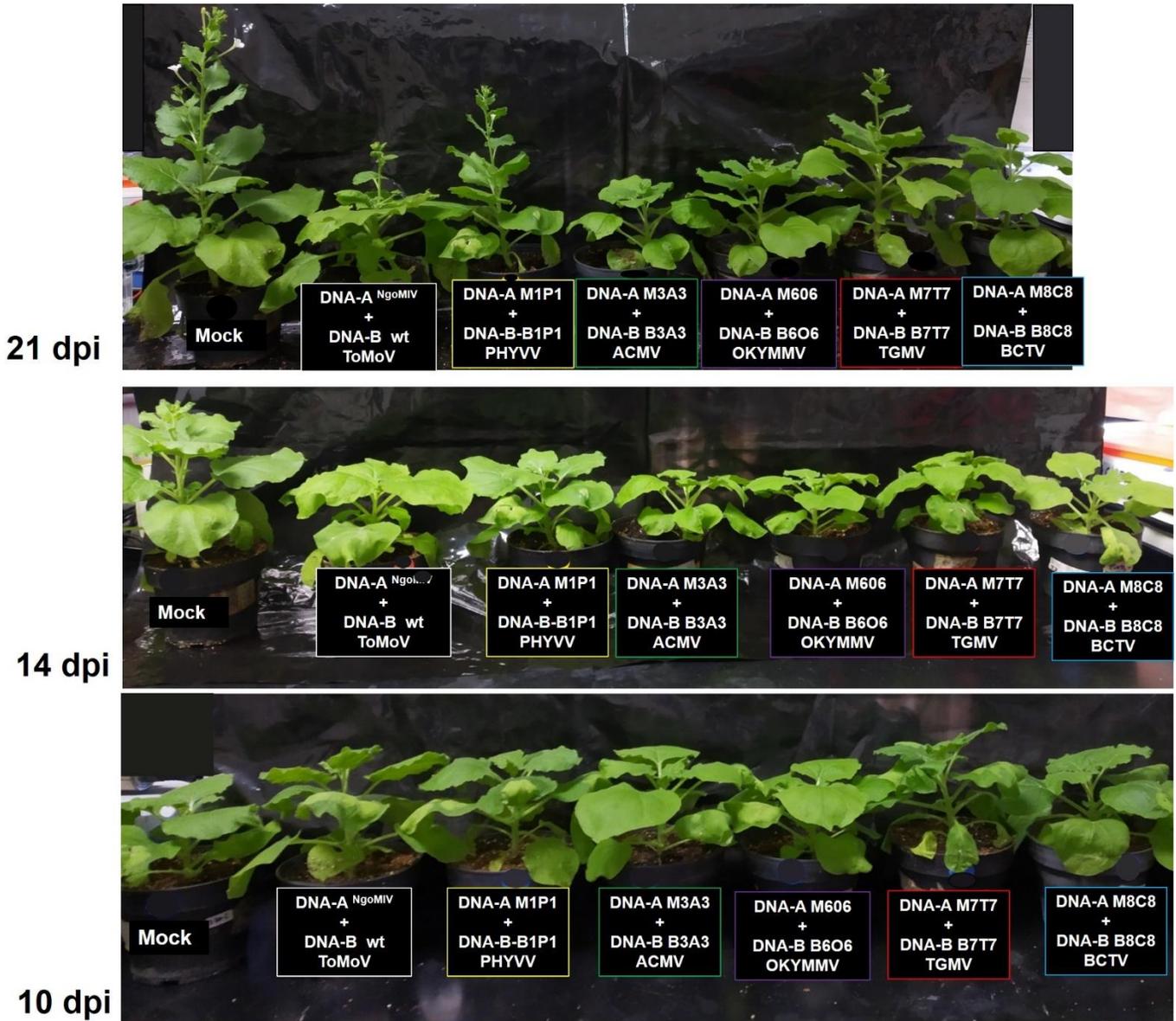


Figura S1- Ilustración de un experimento independiente con ToMoV y los mutantes derivados del mismo con determinantes de especificidad (DEs) de Rep e iterones modificados. En este caso, a diferencia del experimento descrito en las páginas 49-50, las plantas inoculadas tenían 5 semanas de edad al momento de la agroinoculación, por lo que a los 21 dpi algunas plantas infectadas iniciaron la floración, y es probable que la infección empezaba a remitir en algunos casos. Sin embargo, de acuerdo con los datos de PCR (ver Fig. S2), todas ellas tenían DNA viral en las hojas jóvenes (infección sistémica).

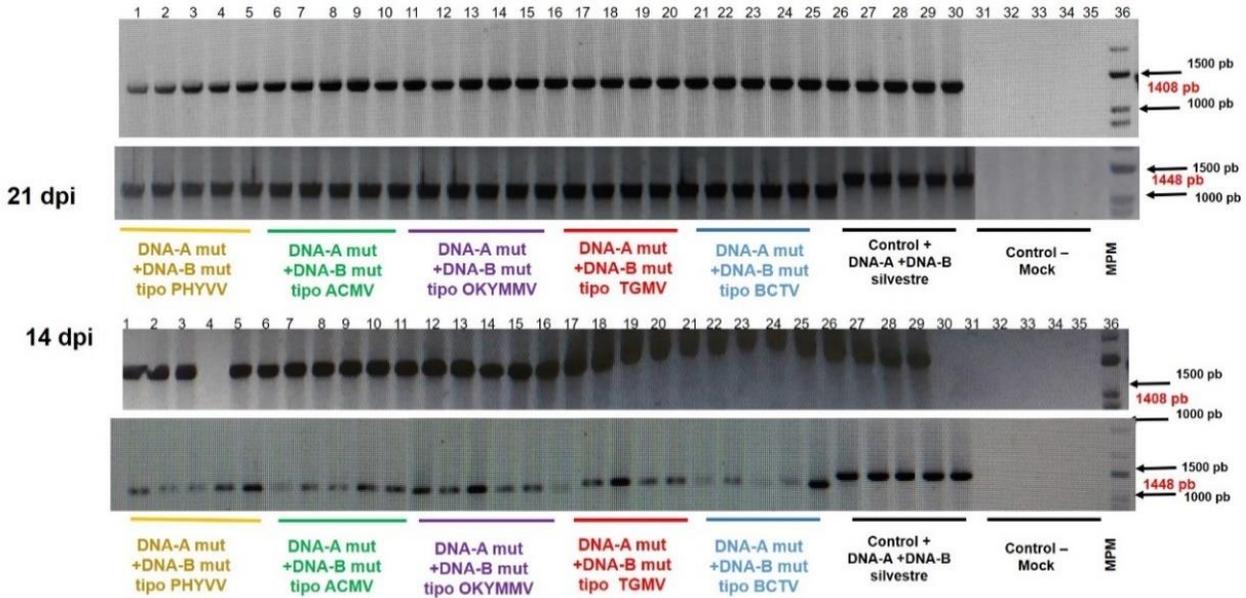


Figura S2. Productos de PCR de los extractos de DNA de 35 plantas del experimento ilustrado en la figura S1. En el caso del gel con los amplicones del DNA-A a los 14 dpi, se observa un mayor tamaño de los productos de PCR a partir del carril 17 y hasta el 29, lo cual no es real sino un artefacto de la electroforesis. En algunos casos, como en los carriles 4 y 30 del gel con las muestras del DNA-A a los 14 dpi no se observa un producto de PCR, pero su ausencia se debió a omisiones involuntarias al cargar las muestras en el gel, como luego lo verificamos. Nótese la diferencia de tamaño de los amplicones de los DNA-B mutantes respecto al DNA-B silvestre.