



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

POSGRADO EN CIENCIAS EN BIOLOGÍA MOLECULAR

**“CARACTERIZACIÓN DE NUEVOS BEGOMOVIRUS DE UN
LINAJE QUE EVOLUCIONÓ POR RECOMBINACIÓN CON UN
GEMINIVIRUS DE ORIGEN DESCONOCIDO”**

Tesis que presenta

SANDRA ILIANA TORRES HERRERA

Para obtener el grado de

Doctor en Ciencias en Biología Molecular

Director de la Tesis

Dr. Gerardo Rafael Argüello Astorga

San Luis Potosí, S.L.P., Marzo del 2015



Constancia de aprobación de la tesis

La tesis “*Caracterización de nuevos begomovirus de un linaje que evolucionó por recombinación con un geminivirus de origen desconocido*” presentada para obtener el Grado de Doctora en Ciencias en Biología Molecular fue elaborada por **Sandra Iliana Torres Herrera** y aprobada el veinte de marzo del dos mil quince 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
Director de la tesis

Dr. Ángel Gabriel Alpuche Solís
Miembro del Comité Tutorial

Dr. Bernardo Bañuelos Hernández
Miembro del Comité Tutorial

Dr. J. Sergio Casas Flores
Miembro del Comité Tutorial

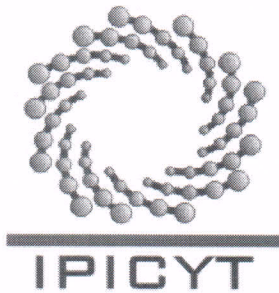


Créditos Institucionales

Esta tesis fué 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 dirección del Dr. Gerardo Rafael Arguello Astorga.

Durante la realización de este trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología CONACYT-70475.

Las investigaciones descritas en esta Tesis fueron financiadas con recursos otorgados al Dr. Gerardo Argüello Astorga por el CONACYT (Proyecto CB-SEP 84004-Q).



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

Acta de Examen de Grado

El Secretario Académico del Instituto Potosino de Investigación Científica y Tecnológica, A.C., certifica que en el Acta 074 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 20 días del mes de marzo del año 2015, se reunió a las 12:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

Dr. Ángel Gabriel Alpuche Solís	Presidente	IPICYT
Dr. J. Sergio Casas Flores	Secretario	IPICYT
Dr. Bernardo Bañuelos Hernández	Sinodal externo	UASLP
Dr. Gerardo Rafael Argüello Astorga	Sinodal	IPICYT

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

DOCTORA EN CIENCIAS EN BIOLOGÍA MOLECULAR

sustentó la C.

Sandra Iliana Torres Herrera

sobre la Tesis intitulada:

Caracterización de nuevos begomovirus de un linaje que evolucionó por recombinación con un geminivirus de origen desconocido

que se desarrolló bajo la dirección de

Dr. Gerardo Rafael Argüello Astorga

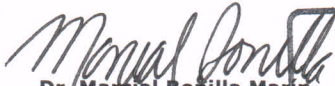
El Jurado, después de deliberar, determinó

APROBARLA

Dándose por terminado el acto a las 13:35 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe el Secretario Académico 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 20 días del mes de marzo de 2015.


Mtra. Ivonne Lizette Cuevas Vélez
Jefa del Departamento del Posgrado


Dr. Marcial Bonilla Marín
Secretario Académico



Dedicatorias

Al Creador de Universo por darme vida.

A mis Ángeles, mis hijos Zadkiel Eduardo y Juan Manuel, por elegirme como madre y por ser el faro y el ancla de mi existencia.

Agradecimientos

Este trabajo se realizó en el laboratorio de Biología Molecular de Plantas de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica. Fue dirigido por el Doctor Gerardo Argüello a quien agradezco por permitirme formar parte de su equipo de trabajo y darme la oportunidad de llevar a cabo este proyecto. Además de mostrar siempre sencillez, calidad humana y mucha paciencia hacia mi persona.

A mis asesores, Dr. Angel Gabriel Alpuche, al Dr. Sergio Casas y al Dr. Bernardo Bañuelos Hernández por el valioso tiempo invertido en la revisión de este trabajo y por sus acertados comentarios. Al Dr. Bernardo le agradezco por sus asesorías tanto en campo como en laboratorio, por su paciencia ante tantas dudas y su disposición para resolvérmelas siempre. Por ser mi compañero en el laboratorio y guiarme a través de mis primeros pasos en el laboratorio de Biología Molecular, y por dejarme convivir con su familia.

A Isabel Isordia “Isabelita” por su apoyo en mis investigaciones, así como enseñarme la técnica de biobalística. A Rosy y Salvador por estar prestos a apoyarme con técnicas del laboratorio. A Janeth que fue un valioso apoyo en el laboratorio y después con mis hijos.

A los Doctores de la División de Biología Molecular que compartieron conmigo sus conocimientos durante las clases impartidas en el programa de maestría, las cuales lleve como oyente, decisión por demás acertada.

A mis amigos y compañeros del Laboratorio de Biología Molecular de Plantas que pusieron un granote de arena para que me puliera y aprendiera de estos nuevos temas para mí.

A mis hijos quienes sacrificaron tiempo de esparcimiento para estar conmigo en las instalaciones del IPY, sábados y domingos sembrando semillas o echándome porras.

A la UJED por permitirme ausentarme de mis responsabilidades para realizar este proyecto personal.

Gracias al CONACYT por la beca otorgada.

Contenido

Constancia de aprobación de la tesis	ii
Créditos institucionales	iii
Acta de examen	iv
Dedicatorias	v
Agradecimientos	vi
Contenido	vii
Lista de figuras	viii
Abreviaturas	x
Resumen	xi
Abstract	xiii
Introducción	1
Materiales y Métodos	11
Resultados	16
Discusión	35
Referencias	43
Anexos	48

Lista de Figuras

Fig.1. Organización genómica de los begomovirus del Nuevo Mundo.	4
Fig.2. Organización del origen de replicación de un begomovirus típico del Nuevo Mundo.	5
Fig.3. Hipotético origen evolutivo de los virus que infectan organismos en los tres dominios principales de la vida.	6
Fig.4. Organización intergénica de los diferentes linajes de geminivirus conocidos en 1994.	9
Fig.5. Plantas con síntomas de virosis de las que se aislaron las cuatro nuevas especies del linaje del SLCV descritas en el estudio	20
Fig.6. Relaciones filogenéticas de los begomovirus aislados en este estudio con un conjunto de BGVs selectos.	22
Fig.7. Evento de recombinación detectado en el DNA-A de AbGMYuV	24
Fig.8. Síntomas inducidos por las clonas generadas de los componentes genómicos A y B de JaCLDYuV (a y d), Cap YSYuV (b y e) y VYMV (c y f).	26
Fig. 9. Motivos en Rep específicos del linaje SLCV.	29
Fig. 10. Filogenia derivada del dominio N- Terminal de Rep	32
Fig. 11. Motivos conservados en las proteínas AC4 de begomovirus del clado SLCV	33
Fig. 12. Filogenia de proteínas AC4/ C4.	34

Fig.13. Alineamiento de la región N- terminal de proteínas Rep de los 30 miembros conocidos del clado SLCV y de los dos curtovirus que poseen proteínas similares. 37

Fig. 14. Dos clases de proteínas Rep caracterizan a los geminivirus de 4 géneros. 39

Fig. 15. Alineamientos de los 12 residuos aminoácidos iniciales de la proteína codificada por el ORF C2 del curtovirus SpiSCTV, con las proteínas homólogas de algunos begomovirus. 41

Abreviaturas

AbMV	Virus del mosaico del abutilón
AbGMV	Virus del mosaico dorado del abutilón
AC2	Gen que codifica la proteína TrAP
AC4	Gen que codifica la proteína AC4
ADN	Ácido desoxirribonucleico
ADNcd	ADN de cadena doble
ADNcs	ADN de cadena sencilla
BCTV	Virus del rizado apical del betabel
BCaMV	<i>Bean calico mosaic virus</i>
BGVs	Begomovirus
CaLCuV	Virus del enchinamiento de la hoja de la col
CpYSYuV	Virus del moteado amarillo de Capraria
CP	Proteína de la cápside
JacLDYuV	Virus de la deformación foliar de Jacquemontia
HrCTV	<i>Horseradish curly top virus</i>
MSV	Virus del rayado del maíz
PCR	Reacción en cadena de la polimerasa
PepGMV	Virus del mosaico dorado del chile
PHYVV	Virus Huasteco del amarillamiento de las venas del chile
pRBR	Proteína relacionada a retinoblastoma
PYMV	Virus del mosaico amarillo de la papa
RCR	Replicación por Círculo Rodante
Rep	Proteína iniciadora de la RCR
SLCV	Virus del rizado foliar de la calabaza
SpiSCTV	<i>Spinach severe curly top virus</i>
TYLCV	Virus del rizado amarillo del tomate
ViYMV	Virus del mosaico amarillo de Vigna

Resumen

Caracterización de nuevos begomovirus de un linaje que evolucionó por recombinación con un geminivirus de origen desconocido.

Los begomovirus (BGVs) son patógenos de plantas dicotiledóneas transmitidos por una sola especie de mosquita blanca, *Bemisia tabaci*, una plaga agrícola cosmopolita. Estos virus evolucionaron en Asia, de donde se diseminaron a otras regiones del mundo, incluyendo el continente americano, al que arribaron durante el Oligoceno según los estimados paleoviroológicos. Los BGVs se dispersaron subsecuentemente por todo el continente, y este proceso de radiación evolutiva dió origen a diferentes linajes secundarios. Uno de éstos, denominado el “clado del *Squash leaf curl virus* (SLCV)” incluye a virus que codifican una proteína iniciadora de la replicación (Rep) cuyo dominio N-terminal difiere de sus homólogos en otros BGVs, además de exhibir en el origen de replicación viral un arreglo único de sitios de unión para Rep (“iterones”). Con el propósito de explorar la diversidad del clado del SLCV en México, y obtener pistas para comprender su evolución, llevamos a cabo una búsqueda sistemática de nuevos miembros de esa estirpe viral en plantas silvestres de la Península de Yucatán y otras regiones de México. En este trabajo describimos cuatro nuevas especies del linaje del SLCV, una de las cuales, *Jacquemontia leaf distortion Yucatan virus* (JaLDYV) es basal al clado, mientras que otra, denominada *Vigna yellow mosaic virus* (ViYMV), exhibe determinantes de especificidad replicativa en *cis* y en *trans* que difieren del resto de los miembros de su linaje. Un análisis comparado de las secuencias codificantes y reguladoras del componente genómico A de los miembros del clado del SLCV y el resto de los BGVs conocidos, reveló que un segmento de ~700 nt, que abarca la región del origen de replicación, el gen AC4 completo y los primeros 160 codones del gen *AC1/rep*, difiere significativamente de la sección genómica equivalente de otros begomovirus. La extensión del análisis comparado a todos los geminivirus descritos al presente reveló la existencia de dos especies de otro género (*Curtovirus*) que poseen un segmento genómico similar al de los miembros del linaje SLCV. Esta observación sugiere la posibilidad de que el ancestro de ese linaje haya surgido de un evento de recombinación entre un curtovirus y un begomovirus. Esta inferencia no fue respaldada por los datos de un análisis más minucioso. La conclusión de nuestro estudio es que el ancestro del linaje del SLCV fue un recombinante derivado de un primitivo BGV del Nuevo Mundo, y un geminivirus de una estirpe hasta ahora desconocida, que debió existir en el continente americano antes de su colonización por los BGVs. Este trabajo proporciona respuestas a cuestiones importantes sobre la evolución de los begomovirus en el Nuevo Mundo, al tiempo que abre nuevas interrogantes sobre la evolución de los geminivirus en el Cretácico, cuando los supercontinentes Gondwana y Laurasia se fragmentaron y dieron origen a los continentes actuales.

Palabras clave: replicación viral, recombinación, círculo rodante, evolución.

Abstract

Characterization of novel begomoviruses belonging to a lineage that evolved by recombination with a geminivirus of unknown origin.

The begomoviruses (BGVs) are highly diversified plant pathogens transmitted by the whitefly *Bemisia tabaci*, a worldwide distributed pest. The BGVs evolved in the Old World and subsequently spread into the Americas engendering several secondary lineages; one of which, named after the *Squash leaf curl virus* (SLCV), includes viral species exhibiting a distinctive arrangement of *cis*-acting elements in the replication origin (*Ori*) region. To further explore the diversity of the SLCV clade and better understand its evolution, we searched for additional members in a variety of wild plants, and looked for molecular features distinguishing to members of this lineage from other New World BGVs. Here we described four novel species of the SLCV clade; one of them (i.e., *Jacquemontia leaf distortion Yucatan virus*) is the most divergent member exhibiting all the lineage distinctive features, whereas a second one (*Vigna yellow mosaic virus*) displayed potential replication specificity determinants that are unique among the BGVs. A systematic comparative analysis of both coding and non-coding sequences of the SLCV clade members revealed that a ~700 bp long genomic segment encompassing the *Ori* region, the entire *AC4* gene, and the first 160 codons of the *AC1/rep* gene, greatly differ from the equivalent genome segment of all other described BGVs. The extension of the analysis to encompass all geminiviruses revealed two curtoviruses displaying a genomic segment similar to that of the SLCV clade. A thorough comparative analysis conclusively demonstrated that the ancestor of the latter lineage was not a recombinant derived from a curtovirus and a begomovirus. Thus, we concluded that the common ancestor of the SLCV clade was a recombinant between an ancient New World BGV and a geminivirus from a hitherto unknown lineage indigenous to the Americas.

Keywords: viral replication, recombination, rolling circle, evolution.

Introducción

Los geminivirus son patógenos de plantas que infectan a una gran variedad de especies cultivadas, incluyendo hortalizas, leguminosas, cereales, plantas productoras de fibras y ornamentales, y causan enfermedades agrícolas severas, muchas veces epidémicas, en todas las regiones del mundo con climas que favorecen el desarrollo de los insectos vectores (Rojas *et al.*, 2005). Los geminivirus se distinguen del resto de los virus conocidos por la estructura de su cápside, que está formada por dos icosahedros incompletos fusionados, lo que da a los viriones la apariencia de partículas gemelas o “geminadas”, cuyo tamaño se ha estimado en 22 x 38 nm (Bottcher *et al.*, 2004; Zhang *et al.*, 2001). Su genoma se compone de una o dos moléculas circulares de ADN de cadena sencilla (ADNcs) de 2500 a 3000 nucleótidos de longitud, las cuales son replicadas en el núcleo de la célula huésped por un mecanismo de círculo rodante (Jeske, 2009). El genoma de estos virus codifica muy pocas proteínas, por lo que dependen extensamente de la maquinaria molecular de la célula para replicar su DNA y expresar sus genes, razón por la cual se han convertido en modelos excelentes para estudiar procesos moleculares básicos de los sistemas vegetales, como el control del ciclo celular, la replicación del ADN, y el transporte sistémico de macromoléculas (Gutiérrez, 2002; Arguello-Astorga *et al.*, 2004). En efecto, diversos estudios han demostrado que los geminivirus reprograman el ciclo celular de las células que infectan para favorecer la replicación del ADN viral, alterando también el patrón de expresión génica y el tráfico macromolecular del huésped; asimismo, interfieren las vías de señalización de la célula y el recambio proteico

para bloquear algunas respuestas de defensa de la planta (revisado por Hanley-Bowdoin *et al.*, 2013).

De acuerdo a datos recientes del Comité Internacional de Taxonomía de Virus (ICTV), la familia *Geminiviridae* constituye la mayor familia viral conocida, superando en número de especies descritas (~300) a todos los grupos de virus que infectan plantas, animales, hongos, algas y otros sistemas eucarióticos o procarióticos, como bacterias y arqueas [www.ictvonline.org/virusTaxonomy.asp].

Los geminivirus son transmitidos por insectos del orden Hemiptera, especializados en la succión de la savia de plantas superiores, y la identidad de sus vectores específicos, así como la organización del genoma viral y la gama de huéspedes que infectan, constituyen los criterios principales para clasificar actualmente a los geminivirus en siete géneros, que son los siguientes: *Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* y *Turncurtovirus* (Adams *et al.*, 2013; Varsani *et al.*, 2014).

El género *Begomovirus* es el más diversificado y ampliamente distribuido a nivel mundial, con más de 280 especies aisladas de plantas dicotiledóneas de todos los continentes, con la excepción de la Antártida. Los miembros de este género poseen un genoma compuesto por uno o dos componentes, y son transmitidos exclusivamente por una especie particular de “mosquita blanca” (Familia *Aleyrodidae*) cuyo nombre científico es *Bemisia tabaci*, la cual es una plaga agrícola cosmopolita muy difícil de controlar (Seal *et al.*, 2006; Rojas *et al.*, 2009). Los begomovirus incluyen algunos de los patógenos que causan mayores pérdidas económicas a la agricultura mundial, entre los que destacan TYLCV (*Tomato yellow leaf curl virus*) que afecta cultivos de tomate, papa, chile y otras

solanáceas; *Cotton leaf curl virus* (CLCV) que infecta plantas de algodón y otras malváceas en Pakistán y la India; y el complejo de especies relacionadas a ACMV (*African cassava mosaic virus*), que causan severas epidemias en cultivos de yuca o cassava (*Manihot esculenta*) en el continente africano, donde esa planta representa la principal fuente de alimento de las poblaciones económicamente desfavorecidas (Legg y Fauquet, 2004; Rojas *et al.*, 2005).

Organización genómica de los begomovirus

Sobre la base de su distribución geográfica y características específicas de su organización genómica, los begomovirus han sido tradicionalmente divididos en dos grandes grupos: los del Viejo Mundo (África, Europa, Asia y Oceanía) y los del Nuevo Mundo (América). Con una excepción recientemente reportada (Melgarejo *et al.*, 2013) todos los begomovirus del continente Americano y una parte significativa (pero no mayoritaria) de los del Viejo Mundo, poseen un genoma compuesto por dos moléculas circulares, denominadas DNA-A y DNA-B (**Fig. 1**). El componente genómico A contiene cinco genes: 1) *AC1* o *rep*, que codifica la proteína iniciadora de la replicación, Rep; 2) *AC2* o *trap*, que codifica una proteína (TrAP) que transactiva la expresión de los genes tardíos y funciona como supresor de la respuesta de silenciamiento; 3) *AC3* o *ren*, que produce una proteína (REn) que optimiza la replicación viral al interactuar con Rep y proteínas reguladoras de la planta; 4) *AC4*, que codifica un supresor del silenciamiento post-transcripcional; y 5) *AV1* o *cp*, que codifica la proteína de la cápside viral (CP). El componente genómico B, por otra parte, incluye dos genes que se transcriben en direcciones opuestas: 1) *BV1* o *nsp*, que codifica una proteína (NSP, Nuclear Shuttle Protein)

que transporta el DNA viral del núcleo al citosol, y también de éste al núcleo; y 2) el gen *BC1* o *mp*, que codifica la proteína de movimiento (MP), involucrada en el movimiento intercelular y sistémico del DNA viral (revisado por Fondong, 2013). La región no codificante que separa a los genes que se transcriben divergentemente, contiene el origen de replicación (*Ori*) viral, localizado en un segmento de entre 150 y 200 nt (dependiendo del linaje begomoviral) que presenta una secuencia altamente similar en ambos componentes genómicos. Ese segmento se denomina “Región Común” y es fundamental para la replicación coordinada de los componentes genómicos del virus (Gutierrez, 2002; Jeske, 2009).

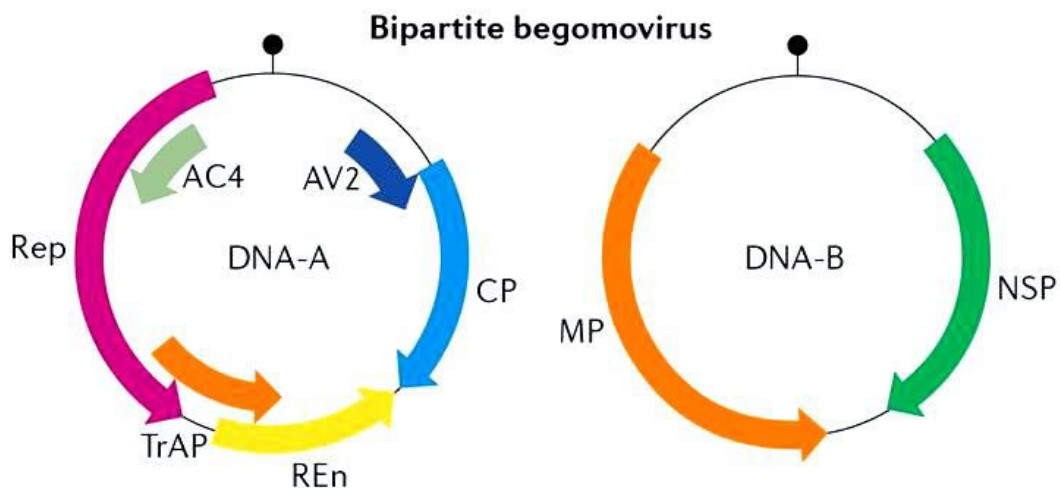


Figura 1. Organización genómica de los begomovirus del Nuevo Mundo (reproducido de Hanley Bowdoin *et al.*, 2013)

Proceso de replicación de los begomovirus

La replicación de los geminivirus se lleva a cabo en el núcleo de la célula huésped. En la primera etapa del proceso se verifica la síntesis de la cadena complementaria del ADNcs genómico, que es catalizada por ADN polimerasas de

la planta, a través de un mecanismo que todavía no se conoce, pero que involucra secuencias presentes en la región intergénica que no han sido delimitadas con precisión (Gutierrez, 2002; Fondong, 2013). Una vez que el ADN circular de doble cadena se ha generado, los genes virales tempranos son transcritos por la ARN polimerasa-II con el concurso de factores transcripcionales de la célula huésped. Uno de esos genes tempranos es *AC1 (rep)*, que codifica la proteína iniciadora de la replicación por Círculo Rodante (RCR) del virus, que interactúa con el origen de replicación viral (*Ori*) a través del reconocimiento de dos tipos diferentes de elementos. El primero de ellos es una secuencia de 9 nucleótidos (5'-TAATATTAC- 3') que se conserva en todos los geminivirus, y se ubica en el centro de un segmento de DNA de ~30 nt con el potencial de adoptar una estructura secundaria de tipo "tallo-asa" (**Fig.2**), donde Rep introduce un corte endonucleolítico ("nick") y genera un extremo 3'-OH libre a partir del cual las ADN polimerasas y otros factores del huésped replican el genoma viral por RCR (Laufs et al, 1995; Singh et al, 2008). El segundo tipo de elemento es un conjunto de secuencias cortas repetidas, localizadas en la vecindad del elemento tallo-asa antes mencionado, a las que la proteína Rep se une con alta afinidad. Estos repetidos se denominan "iterones", y diversos estudios han mostrado que funcionan como determinantes de especificidad replicativa de los geminivirus (Argüello-Astorga *et al.*, 1994a,b; Hanley-Bowdoin *et al.*, 1999; Londoño *et al.*, 2010).

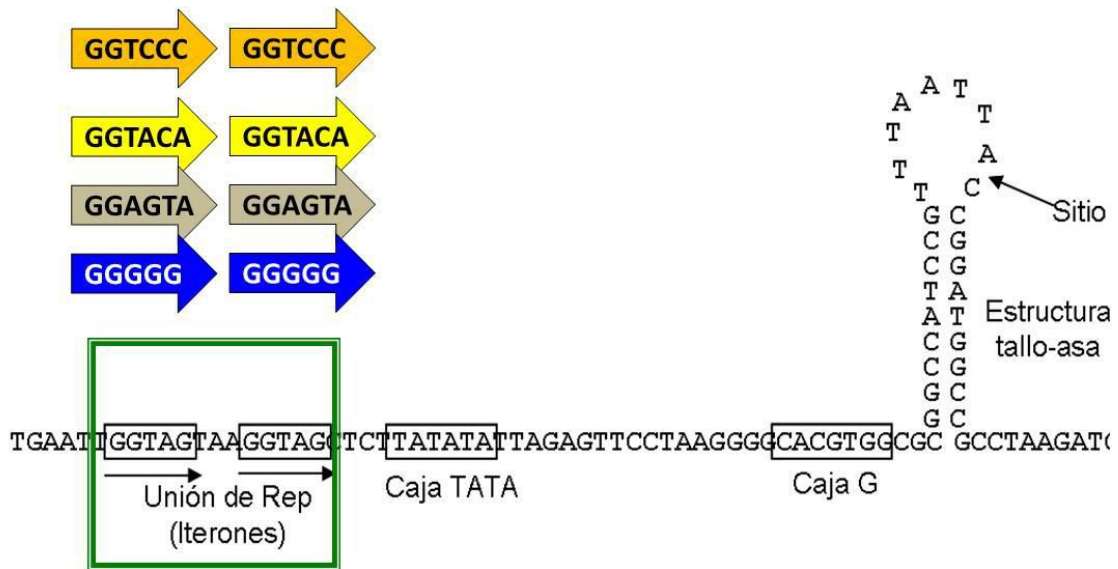


Figura 2. Organización del origen de replicación de un begomovirus típico del Nuevo Mundo. Las secuencias en fondo de color corresponden a iterones de begomovirus diferentes, e ilustran las variaciones en los sitios de unión de Rep (Modificado de Orozco et al., 1998).

La proteína Rep pertenece a un extenso grupo de proteínas iniciadoras de la RCR que se caracterizan por un arreglo de tres motivos aminoácidos conservados, que se denominaron simplemente “Motivos RCR I, II y III” (Ilyina y Koonin, 1992). Estos iniciadores de la RCR, ahora agrupadas en la llamada Superfamilia de Endonucleasas HUH (Chandler *et al.*, 2013), son codificados por una multitud heterogénea de entidades genéticas que incluye a bacteriófagos, plásmidos de bacterias y arqueas, virus de hongos, algas y plantas, plásmidos de algas rojas y protozoarios, así como virus de animales invertebrados y vertebrados, por lo que se considera que esa superfamilia de endonucleasas podría haber evolucionado en el período previo a la divergencia de los dominios primarios Bacteria, Archea, y Eukarya (**Fig.3**) (Iyer *et al.*, 2005; Londoño *et al.*, 2010).

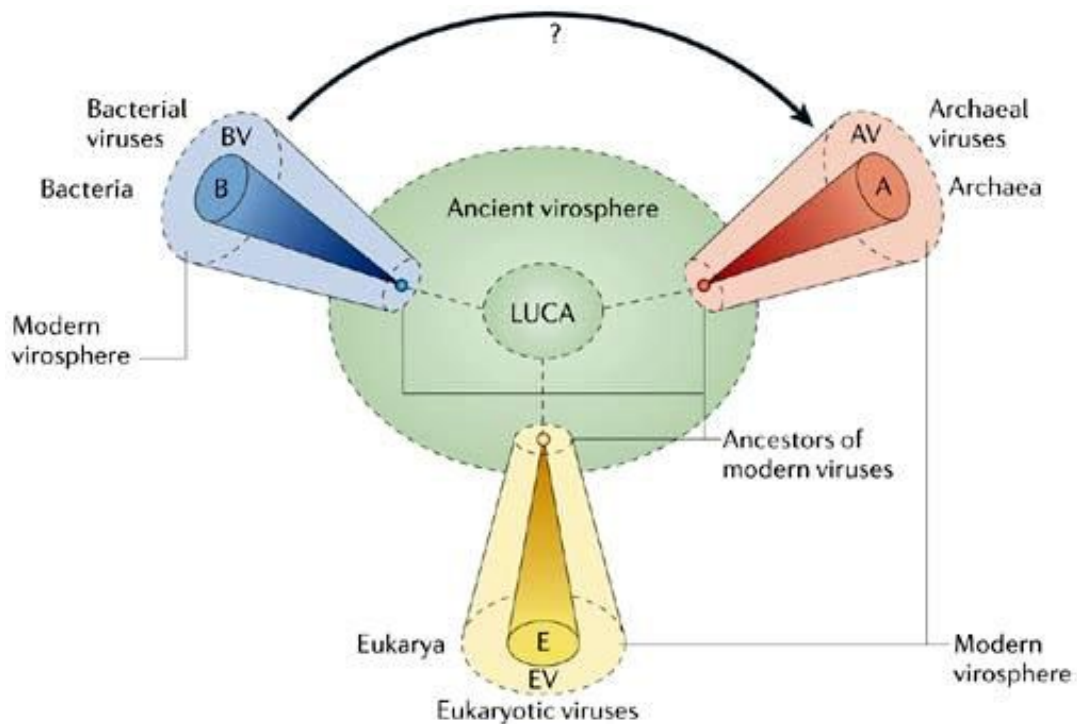


Figura 3- Hipotético origen evolutivo de los virus que infectan organismos en los tres Dominios Primarios de la vida (Reproducido de Prangishvili *et al.*, 2006)

Origen evolutivo de los begomovirus del Nuevo Mundo

Los begomovirus (“BGVs”, por simplicidad) originarios del continente americano se distinguen de sus contrapartes de otros continentes por varias características moleculares, entre las que destacan un número y arreglo de iterones diferente en el origen de replicación (Arguello-Astorga *et al.*, 1994a,b; Arguello-Astorga y Ruiz-Medrano, 2001), la ausencia de un gen (*AV2*) que precede al gen *AV1* (*cp*) en todos los BGVs del Viejo Mundo (VM), y la conservación del motivo PWRsMaGT en el extremo N-terminal de la proteína de la cápside (Harrison *et al.*, 2002; Ha *et al.*, 2006, 2008). No obstante estas diferencias, múltiples líneas de evidencia indican que los BGVs de América evolucionaron a partir de virus del VM (Rybicki, 1994; Nawaz-ul-Rehman y Fauquet, 2009; Briddon *et al.*, 2010). Un estudio reciente que involucró el uso de diversos programas de reconstrucción

filogenética, así como el análisis genómico comparado basado en relojes moleculares, condujo a la plausible conclusión de que los BGVs del NM evolucionaron de BGVs que llegaron al continente americano hace 30-35 millones de años, cuando se formó un puente terrestre entre Asia y Norteamérica. En ese tiempo las condiciones climáticas favorecían la existencia de un cinturón continuo de vegetación subtropical entre los dos continentes (Lefeuvre *et al.*, 2011). De acuerdo a ese escenario hipotético, los primeros BGVs americanos se habrían diseminado de Alaska a todo el bloque continental de Norteamérica durante el Oligoceno, cuando éste último no estaba unido a Sudamérica todavía. Posteriormente, cuando se formó el Istmo de Panamá (~3 millones de años atrás, MYA), la dispersión de mosquitos blancos virulíferos continuó hasta las regiones más meridionales de América con climas permisivos a su proliferación. En ese proceso de radiación evolutiva se originaron varios linajes secundarios o “clados” de BGVs a lo largo del continente americano. Uno de estos linajes, denominado el “clado del *Squash leaf curl virus* (SLCV)”, por ser este último el miembro del grupo primeramente caracterizado (Lazarowitz y Lazdin, 1991), incluye un número importante de BGVs que se distribuyen desde el sur de Estados Unidos hasta Uruguay y Argentina (Bridson *et al.*, 2010; Brown *et al.*, 1999, 2002; Gregorio-Jorge *et al.*, 2010; Marquez-Martin *et al.*, 2012).

Características distintivas de begomovirus del clado del SLCV

A diferencia de otras ramas evolutivas de los BGVs del continente americano, como el clado del AbMV y el “linaje Brasileño”, que fueron definidos en diversas reconstrucciones filogenéticas en base a las secuencias de los DNA-As de un

número significativo de miembros, los virus del clado del SLCV fueron diferenciados de otros BGVs cuando sólo se habían caracterizado 3 especies del linaje (SLCV, PepGMV y BCaMV) basándose en ciertas características atípicas de su origen de replicación, como el número y arreglo peculiar de los iterones, y la ausencia de un motivo CACGTGG (“caja G”) adyacente al elemento tallo-asa, que se conserva en otros BGVs del continente americano (Argüello-Astorga *et al.*, 1994a,b). Adicionalmente, los iterones de los miembros del linaje del SLCV exhiben una secuencia consenso que no parece derivarse de la misma secuencia ancestral de los elementos equivalentes en otros BGVs del NM y el VM (**Fig. 4**) (Argüello-Astorga *et al.*, 1994a, b). Conforme se fueron describiendo más especies del clado SLCV, se hizo claro que también la proteína Rep codificada por estos virus posee un dominio amino-terminal que difiere, significativamente, del que exhiben las proteínas homólogas de otros BGVs (Argüello-Astorga y Ruiz-Medrano, 2001; Gregorio-Jorge *et al.*, 2010). Estas diferencias de los miembros del clado del SLCV respecto a otros BGVs del mismo continente, con los cuales tienen una incuestionable relación filogenética, plantean varias interrogantes evolutivas muy interesantes, que no han sido investigadas de modo sistemático hasta ahora, a pesar de su evidente importancia para comprender mejor la historia evolutiva de los begomovirus en el Nuevo Mundo.

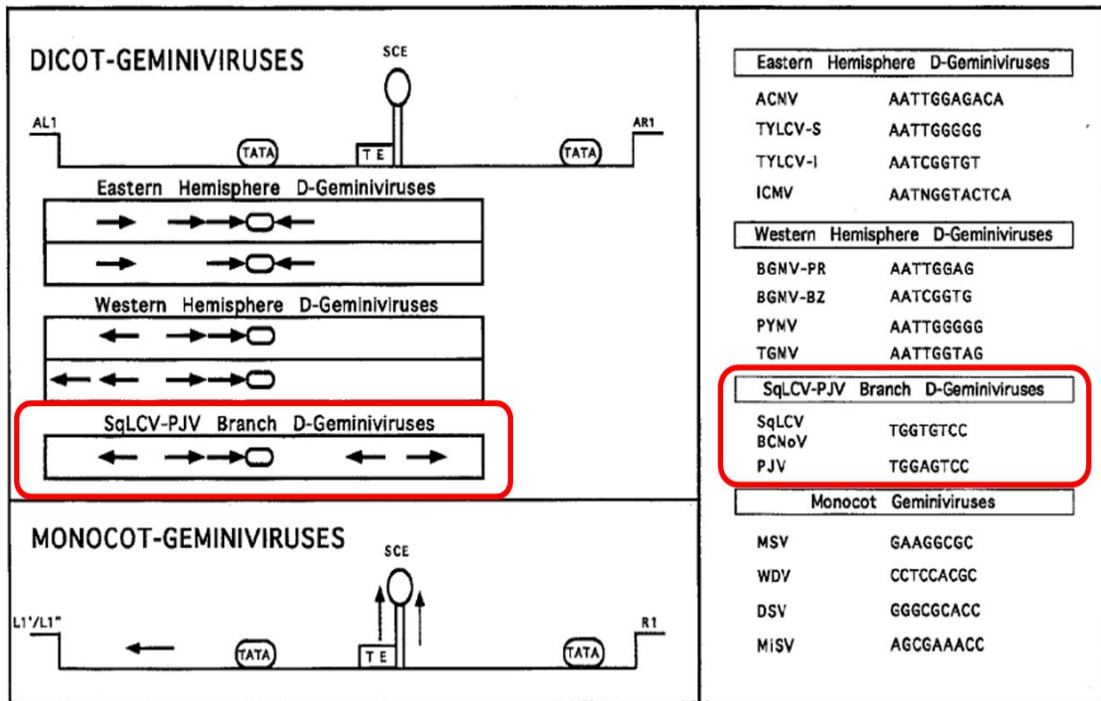


Figura 4- Organización de la región intergénica de los diferentes linajes de geminivirus conocidos en 1994. Las flechas representan los iterones, cuyo número y arreglo son característicos de esos linajes. El clado SLCV es definido aquí como “la rama de SqLCV-PJV” (PJV fue luego denominado PepGMV). Se destacan también las diferencias en secuencia de los iterones de esos virus, cuyo “nucleo” GGNGN no está precedido por el consenso AATY, como en el resto de los BGVs (reproducido de Argüello-Astorga *et al.*, 1994b).

Objetivos

Objetivo general

Reconstruir la historia evolutiva de los begomovirus del clado del SLCV utilizando enfoques múltiples pero complementarios, como el aislamiento y caracterización molecular de nuevas especies de ese linaje, la generación de filogenias a partir de componentes genómicos y proteínas específicas, y la comparación sistemática de proteínas y secuencias no codificantes de esos virus.

Objetivos específicos

- 1- Aislar y caracterizar nuevas especies de begomovirus del linaje del SLCV que infecten plantas silvestres pertenecientes a diferentes familias de dicotiledóneas.

- 2- Generar clonas infectivas de los begomovirus aislados y producir infecciones en hospederos específicos.

- 3- Definir, a través de análisis comparativos exhaustivos, los rasgos moleculares que distinguen a los virus del clado del SLCV de otros begomovirus del continente americano.

- 4- Utilizar la información recabada del genoma de las nuevas especies y sus parientes para, mediante el análisis filogenético y comparativo de secuencias codificantes y reguladoras, encontrar indicios para reconstruir la historia evolutiva del linaje del SLCV.

Materiales y Métodos

Muestras vegetales y extracción del DNA

Se tomaron hojas jóvenes de plantas silvestres que exhibían síntomas de virosis, colectadas en ecosistemas naturales con poblaciones de mosquita blanca. Los síntomas incluyeron deformaciones o enchinamientos foliares, mosaicos amarillos o dorados, clorosis en las nervaduras, y/o retardo en el crecimiento (enanismo). Los sitios de colecta fueron diversos e incluyeron localidades de Yucatán, Campeche y Morelos. Se prepararon extractos de DNA total de las muestras colectadas utilizando un protocolo modificado de Dellaporta *et al.* (1983).

Reacción en cadena de la Polimerasa (PCR)

La amplificación por PCR se llevó a cabo empleando oligonucleótidos degenerados que hibridan con secuencias conservadas de los componentes genómicos A y B de begomovirus del clado SLCV, que son descritos en Gregorio-Jorge *et al.*, 2010. La amplificación con esos iniciadores permite obtener la secuencia completa de cada componente genómico con solo dos pares de iniciadores cada uno, ya que los amplicones que se producen cubren la totalidad de las moléculas del ADN que componen el genoma begomoviral. Las parejas de oligonucleótidos empleados en la amplificación del ADN-A fueron las siguientes: YMAC For/ DGRS Rev, YMAC Rev/ DGRS For (linajes de begomovirus “típicos”), y SL 2150 For/YMAC Rev, SL2200 Rev/ YMAC For (linaje SLCV). En el caso del componente genómico B se utilizaron los iniciadores BV1 310 For/BC1 290Rev y

BV1 310Rev/ BC1290For (Mauricio-Castillo, 2011). Se prepararon mezclas de PCR que contenían la siguiente composición: Buffer Taq DNA polimerasa 1X, MgCl₂ 1.5 mM, dNTPs 0.2mM, oligonucleótidos 1μM y 1μg de DNA. Las reacciones se llevaron a cabo bajo las siguientes condiciones de amplificación: desnaturalización inicial a 94° /2 min, 38 ciclos de 94°C/1 min, 55°C/1 min, 72°C/1 min, con una extensión final de 72°C/5 min. Finalmente los productos obtenidos fueron analizados por movilidad electroforética en geles de agarosa al 1.0%.

Clonación y secuenciación

Los productos de PCR obtenidos se clonaron directamente en el plásmido pGEM-Teasy (Promega), según las indicaciones del proveedor. Posteriormente se realizó la transformación en células competentes de *Escherichia coli* Top 10 por choque térmico con el empleo de cloruro de calcio (CaCl₂), facilitando la selección de transformantes por color blancas o azules, de colonias crecidas en medio LB conteniendo ampicilina, IPTG y X-gal. La extracción del ADN plasmídico se realizó de acuerdo al procedimiento de Birnboim (Sambrook y Rusell, 2001). Las clonas de productos de PCR fueron sometidos a análisis por RFLPs (Restriction Fragment Length Polymorphisms) utilizando las enzimas *Eco RI* y *Hinf I* (New England's BioLabs), para distinguir virus diferentes. Los productos digeridos se analizaron por su movilidad electroforética en geles de agarosa al 2.0%. Los amplicones clonados fueron secuenciados en la Unidad de Secuenciación del CINVESTAV, Irapuato, y las secuencias nucleotídicas obtenidas se compararon con las secuencias disponibles en la base de datos del Gen Bank, utilizando el programa BLAST y los algoritmos ClustalV y ClustalW (MegAlign, DNASTAR).

Generación de clonas infectivas

Para la obtención de las clonas de los componentes genómicos A y B completos de los virus CapYSYuV, JacLDYuV, AbGMYuV, y VMYV, el ADN total extraído de las plantas de *Capraria biflora*, *Jaquemontia penthanta*, *Abutilon permolle* y *Vigna elegans* colectadas fue amplificado por RCA (Rolling-Circle Amplification) usando el kit TempliPhi (GE Healthcare, USA) de acuerdo a las recomendaciones del proveedor. Los componentes A completos de los virus fueron obtenidos cortando el DNA amplificado por RCA con enzimas de corte único: VMYV con *BamHI*, CapYSYuV con *Bstx I*, JaDLYuV con *BamHI* y AbMYuV con *BamHI*, posteriormente los amplicones obtenidos fueron clonados en un vector plasmídico. Se diseñaron oligonucleótidos específicos para usarlos como iniciadores divergentes (“cola a cola”) en sitios de restricción únicos de los componentes genómicos de los cuatro begomovirus. El ADN amplificado por RCA se utilizó para realizar una PCR estándar con los oligonucleótidos divergentes. En el caso de la planta de *Jacquemontia penthanta* se emplearon los siguientes oligonucleótidos: JacA-Bam HI-For (ACT GGA TCC CCT GCC TTA ATG ATT GC) y JacA- Bam HI Rev (TTA GGA TCC AGT AAG ATG GAT AAC C) para amplificar el ADN-A, y JacB-Hind I-For (TAT AAG CTT TCT CTC CCA TCT CCC GT) y JacB- Hind I-Rev (AGA AAG CTT ATC AGA TGC GCA TCC ATG) para amplificar el ADN-B. Los amplicones de ~2.6 Kb fueron clonados en el vector pGEM-Teasy (PROMEGA) y secuenciados. Se procedió de manera análoga con el extracto de la muestra *Capraria-1055*, empleando los siguientes iniciadores: CapA-Bst XI-For

(TAC CCA TCG TGT TGG TAA GCG TTT CTG C) y CapA- Bst XI- Rev (TTA CCA ACA CGA TGG GTA ATG CCA TTG C) para amplificar el ADN-A, y CapB-Sal I-For (ACT GTC GAC AGC TTT AAC TTT CCT TTG) y CapB-Sal I-Rev (TCT GTC GAC AGC TAA GCA CTC TGT AG) para obtener el componente B. En el caso del virus aislado de *Abutilon permolle* se emplearon los oligonucleótidos AbuA-BamHI-For (TAG GAT CCC TCC TTT AAT TTG ACC TGG TCT GCC) y AbuA-BamHI-Rev (CTG GAT CCC ATC AAT CGT GCT GCT TTG CAA TCC) específicos del componente A. El virus aislado de Vigna fue amplificado con los iniciadores VignaA-BamHI-For (ATG GAT CCT CTA TCC TGT ATA TC) y VignaA-BamHI-Rev (TCA TAT GGT CTA CGG TCT CTG ACT AAC) específicos del componente A, y VignaB-KpnI-For (TAG GTA CCG TCA AGA TTG AGC GTG TAC) y VignaB-KpnI-Rev (TTG GTA CCC TTG AAA CGG AGT CG) específicos del componente B. Ambos componentes fueron clonados en el vector pGem-Teasy (PROMEGA).

Inoculación por biobalística.-

Se utilizaron como inoculos las clonas infectivas de CapYSYuV, JacLDYuV, y VMYV, y como control el vector vacío (pGEM-Teasy. Se realizó una preparación de partículas de tungsteno con los diferentes ADNs virales y no-virales según el protocolo de Pons *et al.* (2001); las clonas con los componentes virales fueron digeridas previamente con la enzima de restricción apropiada. Las plantas de *Nicotiana benthamiana* se inocularon aproximadamente al mes de germinadas, cuando presentaban 3-4 hojas verdaderas. Para la inoculación por biobalística se utilizó el dispositivo PDS-1000/He System (BioRad) con una presión de 650 psi.

Análisis filogenéticos

Las secuencias completas de los begomovirus JacLDYuV, CapYSYuV, AbGMYuV, y VMYV fueron comparadas con secuencias de begomovirus del Nuevo y Viejo mundo, disponibles en la base de datos del GeneBank, usando el programa BlastN. La posición y los tamaños de los marcos de lectura abiertos para cada uno de los virus fueron predichos usando el programa EditSeq de DNASTAR, Madison, WI. Los alineamientos fueron obtenidos utilizando el método Clustal W en la aplicación del MegAlign de DNASTAR, usando los parámetros disponibles por default. Los árboles filogenéticos de los componentes A y B de los virus identificados fueron construidos utilizando Mega 4.0.

Resultados

Análisis de plantas silvestres con síntomas de virosis

Durante el año de 2007 el grupo de Virología Vegetal del Centro de Investigaciones Científicas de Yucatán (CICY), dirigido por el Dr. Oscar Moreno, realizó un amplio muestreo de plantas silvestres con síntomas de virosis en diversas regiones de Yucatán, Campeche y Morelos, como parte de un extenso estudio encaminado a explorar la diversidad de BGVs en ecosistemas naturales de México. Se colectaron malezas pertenecientes a familias de dicotiledóneas muy diversas (*Acanthaceae*, *Convolvulaceae*, *Nyctaginaceae*, *Scrophulariaceae*, y otras) con la intención de aumentar las probabilidades de encontrar especies begomovirales no descritas, ya que la mayor parte de las investigaciones en esta área se han centrado en grupos de plantas que incluyen especies de valor agrícola, como las familias *Solanaceae*, *Malvaceae* y *Fabaceae*. Se obtuvieron extractos de ADN total de las muestras vegetales y se procedió a determinar la presencia de BGVs amplificando por PCR un segmento de ~560 pb del gen que codifica la proteína de la cápside (CP) de esos virus utilizando los iniciadores diseñados por Wyatt y Brown (1996). Por un acuerdo de colaboración entre los laboratorios de Virología del CICY y el IPICYT, los extractos de ADN total de doce de las muestras de campo que resultaron positivas para begomovirus fueron enviadas a nuestro laboratorio, para llevar a cabo la caracterización genómica más amplia de los BGVs presentes en las mismas, mediante la técnica LISOP-RFLP desarrollada en el IPICYT. En esta última metodología se utilizan oligonucleótidos sintéticos (“iniciadores”) específicos de linaje viral, los cuales hibridan con

secuencias particulares del gen *AC1 (Rep)* que no comparten los miembros del clado SLCV con otros begomovirus del NM y VM, razón por la cual se definió el método como “PCR traslapante específico de linaje” (LISOP, por sus siglas en inglés). Las diversas combinaciones de “iniciadores SL” e iniciadores universales del ADN-A begomoviral, hacen factible la amplificación específica de virus del clado del SLCV, en tanto que las combinaciones de “iniciadores DGR” y universales del ADN-A amplifican el componente genómico A de BGVs típicos de América y otros continentes (Mauricio-Castillo, 2006; Gregorio-Jorge *et al.*, 2010). Por otra parte, la amplificación del ADN- B de los BGVs de todas las estirpes del Nuevo Mundo, incluyendo la del clado SLCV, se lleva a cabo con iniciadores degenerados comunes, pues no existen secuencias conservadas distintivas de linaje en esos componentes genómicos.

La secuenciación de los productos de PCR obtenidos a partir de las muestras antes mencionadas, reveló una colección de especies virales que ya habían sido reportadas en la Península de Yucatán y/o en otras regiones del país. Tres de las especies virales previamente descritas pertenecen al clado del SLCV: *Rhynchosia golden mosaic Yucatan virus* (RhGMYuV), *Euphorbia mosaic virus-YP* (EuMV-YP), y *Pepper golden mosaic virus* (PepGMV). Adicionalmente, se identificaron BGVs de otros linajes del NM, como *Corchorus yellow spot virus* (CoYSV), *Pepper huasteco yellow vein virus* (PHYVV), *Boerhavia yellow spot virus* (BoYSV), *Sida yellow mosaic Yucatan virus* (SiYMYuV), y *Bean golden yellow mosaic virus* (BGYMV) (datos no publicados). Aparte de los virus enlistados, se identificaron otros 4 amplificados con los iniciadores SL que, por los datos de secuenciación parcial, no parecían pertenecer a especies previamente descritas, por lo que

procedimos a caracterizar su genoma completo. Un dato interesante derivado del análisis de las 12 muestras mencionadas es que 9 de las plantas sintomáticas estaban infectadas por dos BGVs diferentes, pertenecientes a distintos linajes, con una excepción.

Caracterización molecular de nuevas especies de begomovirus del Clado SLCV

Cuatro de las muestras de campo que contenían dos virus diferentes (YP#122, YP#446, YP#534 y Mor# 991) produjeron, con los iniciadores específicos del clado SLCV, amplicones cuya secuencia nucleotídica mostró una divergencia significativa de otros BGVs en las bases de datos públicas, por lo que procedimos a obtener la secuencia completa de sus componentes genómicos. A partir de productos de PCR traslapados (ver Materiales y Métodos) se obtuvo la secuencia del ADN-A y el ADN-B de tres de esos BGVs, pero del cuarto virus solo se logró aislar su ADN-A, pese a repetidos intentos por obtener el segundo componente de ese BGV, el cual probablemente no estaba presente en la muestra. Las secuencias completas de los DNA-As de los 4 virus se sometieron a un análisis con BLASTn para establecer su identidad global con otras especies de BGVs, y determinar si los valores de identidad con su pariente más cercano era, en efecto, inferior al límite establecido por el ICTV (<89%) para definir especies separadas dentro del género *Begomovirus* (Fauquet et al., 2008). El análisis mostró que los virus aislados representan, en efecto, nuevas especies de BGVs. El más divergente de esos virus fue aislado de una planta de *Jacquemontia pentantha* (Convolvulaceae) que mostraba distorsión de las hojas, sin mosaicos

(**Fig.5a**; muestra YP#534), y que fue colectada en el Jardín Botánico de Mérida, Yuc. (coordenadas 21°01'42"N, 89°38'17"W). El ADN-A de este virus, provisionalmente denominado *Jacquemontia leaf distortion Yucatan virus* (JacLDYuV), consta de 2602 nt y mostró la mayor identidad (69.5%) con dos especies no relacionadas, *Tomato severe leaf curl virus* (ToSLCV) de México (No. de acceso: NC_004642) y *Sida yellow leaf curl virus* from Brazil (No. de acceso: EU710750). El componente B tiene 2591 nt de longitud, y comparte un segmento de 190 nt (96% de identidad) con el ADN-A cognado.

Una planta de *Capraria biflora* (Scrophulariaceae) que exhibía acortamiento de los internodos y hojas con moteados amarillos (**Fig.5b**; muestra YP#446), fue colectada en Conkal, Yucatán (coordenadas: 21°06'56"N; 89°53'33"W) y de ella se aisló un virus

cuyos componentes A y B constan de 2624 y 2584 nucleótidos, respectivamente. Su región común comprende solo 154 nt, y presenta una identidad muy baja entre componentes, de solo 72%, aunque los elementos relevantes para la replicación (iterones, elemento tallo-asa) se encuentran bien conservados. El pariente más cercano de este BGV, que denominamos *Capraria yellow spot Yucatan virus* (CapYSYuV), es *Rhynchosia mosaic Sinaloa virus* (DQ406672), con el que comparte una identidad en secuencia global de 81%. Un tercer nuevo virus fue obtenido de una planta de *Abutilon permolle* (Malvaceae) que mostraba mosaicos en las hojas (**Fig.5c**; muestra YP#122), y fue colectada en Suytunchen, una localidad cercana a Mérida, Yuc. (21°01'08" N; 89°53'13" W). El ADN-A de *Abutilon golden mosaic Yucatan virus* (AbGMYuV) es de 2629 nt, y exhibe la mayor similitud en secuencia (81%) con *Desmodium leaf distortion virus*

(DQ875870), también aislado en Yucatán. El cuarto virus se aisló de una planta de *Vigna elegans* con síntomas de virosis consistentes en mosaicos amarillos y hojas ligeramente distorsionadas (**Fig.5d**, muestra Mor# 991), que se colectó en Yauhtepec, Morelos (18° 50'17''N; 98°59'03''W). El ADN-A de esta nueva especie, *Vigna yellow mosaic virus* (ViYMV), consta de 2602 nt y presenta la mayor identidad en secuencia (85%) con el componente genómico homólogo de *Bean calico mosaic virus* (No. acceso: AF110189) de Sonora, Mexico; la molécula del ADN-B tiene 2564 nt de longitud, y presenta un segmento de 192 nt casi idéntico en secuencia (SI=97%) con su DNA-A cognado.

Como se mencionó antes, las 4 nuevas especies del clado SLCV descritas fueron aisladas de plantas que estaban infectadas por dos BGVs, y la identidad del virus co-infectante fue establecida a través de la secuenciación de amplicones de ambos componentes genómicos. Estos virus co-infectantes fueron los siguientes: BYGMV, en el caso de *Vigna elegans*; PHYVV, en el de *Jacquemontia penthanta*; y CoYSV en los casos de *Capraria biflora* y *Abutilon permolle* (datos no mostrados).

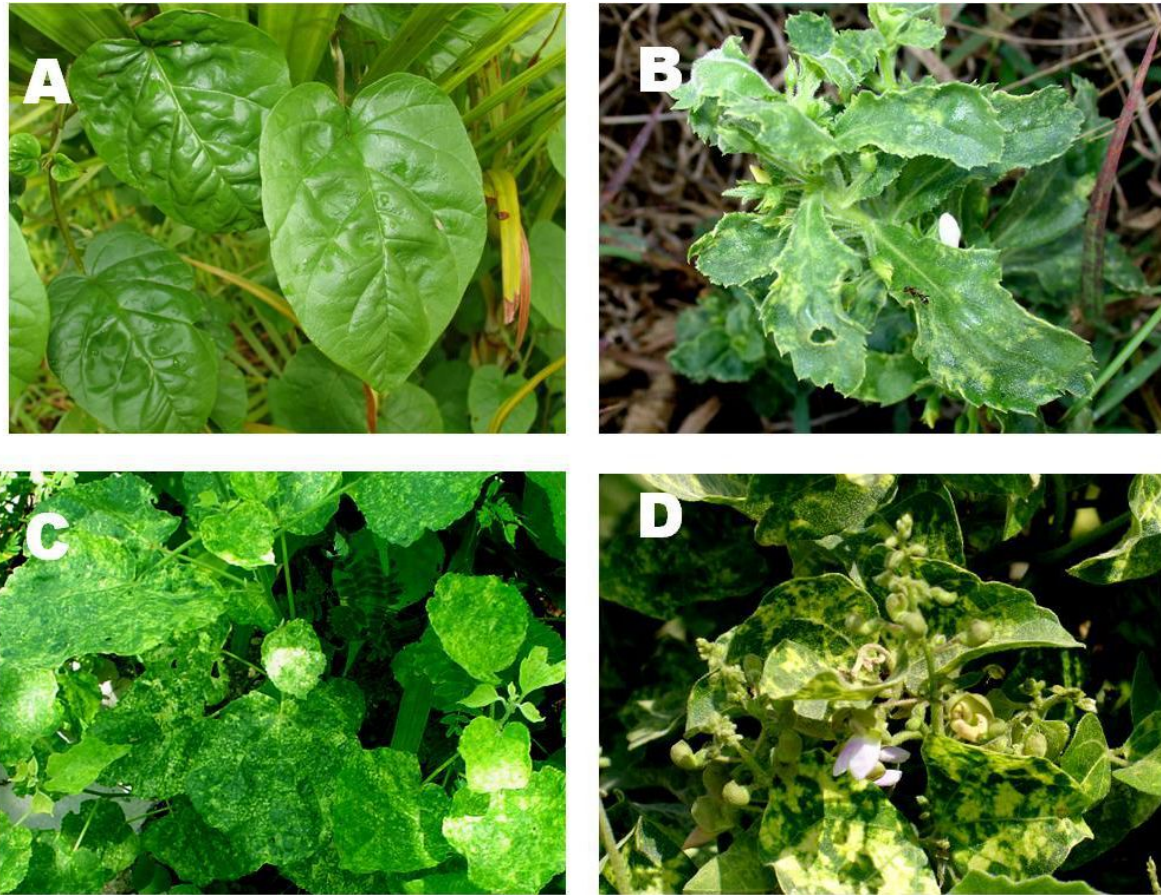


Figura 5- Plantas con síntomas de virosis de las que se aislaron las cuatro nuevas especies del linaje SLCV descritas en este estudio. A- *Jacquemontia penthanta*. B- *Capraria biflora*. C- *Abutilon permolle*. D- *Vigna elegans*.

Relaciones filogenéticas de los begomovirus aislados

Con las secuencias completas de los ADN-As de los nuevos begomovirus y cerca de 45 BGVs con genomas bipartitas, tanto del continente americano como del VM, generamos un árbol filogenético usando el método “neighbor-joining” con 1,000 repeticiones “bootstraps”; el árbol fue enraizado con la secuencia de *Maize streak virus* (MSV), la especie tipo del género *Mastrevirus*. El dendograma resultante (**Fig. 6a**) confirmó la pertenencia de los begomovirus caracterizados al linaje del SLCV. Así, AbGMYuV se agrupó en una rama secundaria que contiene tanto a

BGVs mesoamericanos como sudamericanos, mientras que VYMV y CapYSYuV quedaron incluidos en un diferente grupo de BGVs mesoamericanos. La localización de JacLDYuV en el dendrograma resultó muy interesante, ya que fue basal al linaje del SLCV. Por otra parte, la reconstrucción filogenética derivada de los componentes genómicos B no fue del todo congruente con el dendrograma derivado de la comparación de los ADN-As. Por ejemplo, el ADN-B de JacLDYuV quedó incluido en el clado de AbMV, en tanto que el ADN-B de CapYSYuV DNA-B mostró estar más relacionado con los componentes homólogos del grupo de PHYVV que con los de miembros del clado SLCV (**Fig. 6b**). La incongruencia de las filogenias derivadas de los componentes genómicos A y B, que ha sido observada en numerosos estudios, reflejan la frecuencia relativamente elevada de eventos de recombinación intermolecular y pseudo-recombinación (intercambio de componentes genómicos) en la historia evolutiva de esas especies virales (Bañuelos-Hernández *et al.*, 2012; Briddon *et al.*, 2010; Brown *et al.*, 2004).

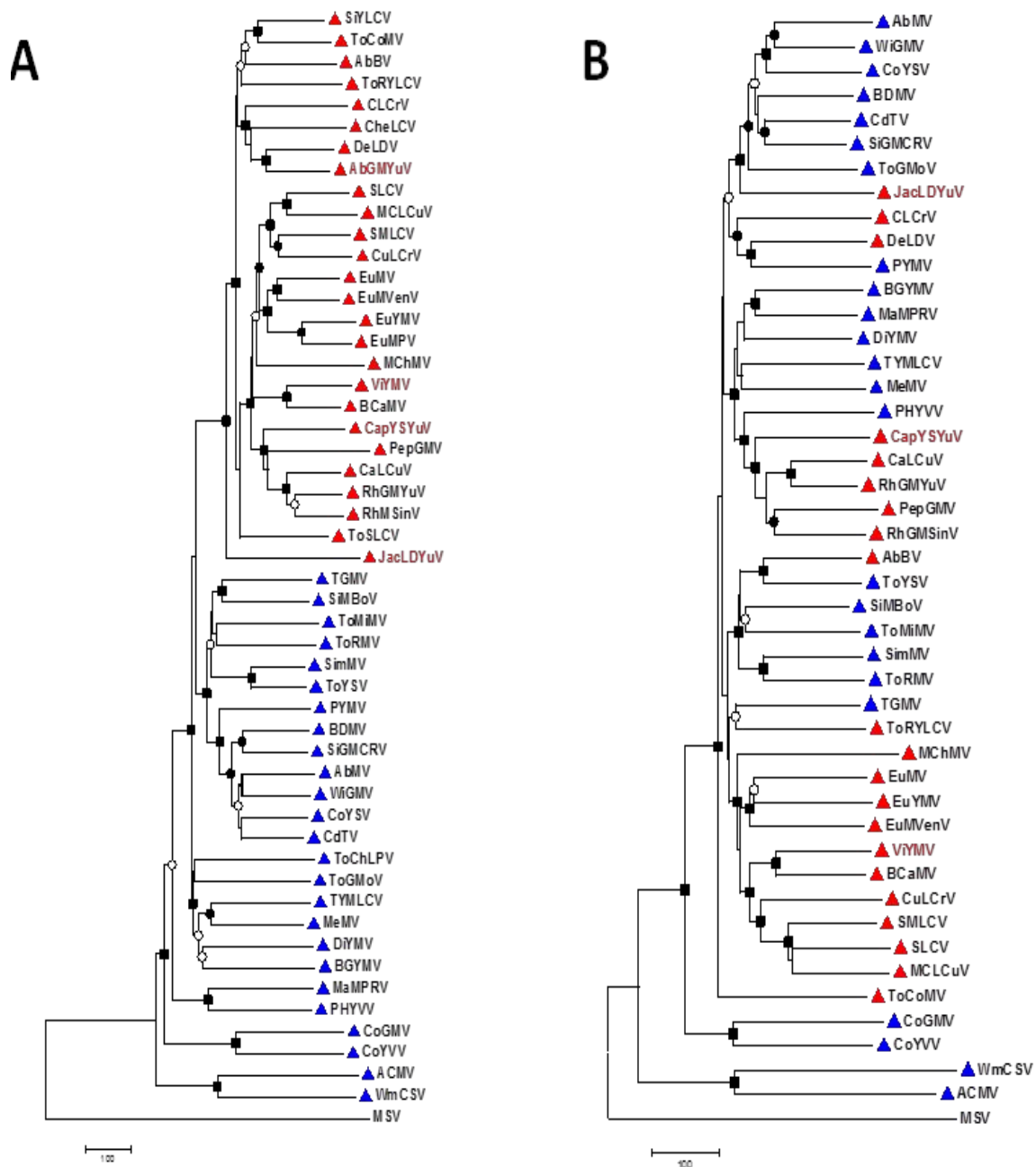


Figura 6. Relaciones filogenéticas de los begomovirus aislados en este estudio con un conjunto de BGVs selectos. Arbol filogenético derivado de los componentes genómicos A y B, respectivamente. Los miembros del linaje del SLCV se indican con un triángulo rojo, en tanto que los begomovirus de otros linajes son denotados con un triángulo azul. El mastrevirus MSV fue utilizado como elemento externo. El análisis filogenético fue realizado con MEGA-5.

Análisis de Recombinación

La secuencia del DNA-A de los 4 nuevos BGVs fue analizada con el conjunto de programas para detectar eventos de recombinación que incluye el paquete RDP4 (Martin *et al.*, 2010). Este análisis no reveló evidencia significativa de recombinación en los casos de CapYSYuV, JacLDYuV, y ViYMV (datos no mostrados), pero sí en el de AbGMYuV. En efecto, en este BGV se detectó un evento de recombinación que involucra al segmento comprendido entre las posiciones 1969 y 2595 del DNA-A. Este presunto segmento recombinante de 626 pb se derivó, aparentemente, de un virus relacionado a *Desmodium leaf distortion virus* (DeLDV), que fue aislado en la misma región geográfica (Yucatán), y que sería el “progenitor menor” de AbGMYuV. La presunta región recombinante abarca la totalidad del gen AC4, la mitad 5´ del gen AC1, y parte de la Región Común del DNA-A (**Fig. 7**). La restante parte de éste último componente genómico se habría derivado de un virus cercano a CoYSV, que sería el “progenitor mayor”. El evento de recombinación entre CoYSV y DesLDV fue estadísticamente sustentado por 6 métodos del paquete RDP4, que fueron los siguientes: GENECONV, $p=2.649 \times 10^{-34}$; RDP, $p=3.606 \times 10^{-31}$; BootScan, $P=6.429 \times 10^{-27}$; MaxChi, $p=1.172 \times 10^{-27}$; Chimaera, $p=1.636 \times 10^{-25}$; y 3Seq, $p=7.617 \times 10^{-50}$. Es oportuno señalar aquí que el hipotético progenitor mayor de AbGMYuV (esto es, CoYSV) no pertenece al clado del SLCV, de acuerdo a la filogenia ilustrada en la **Fig.6**.

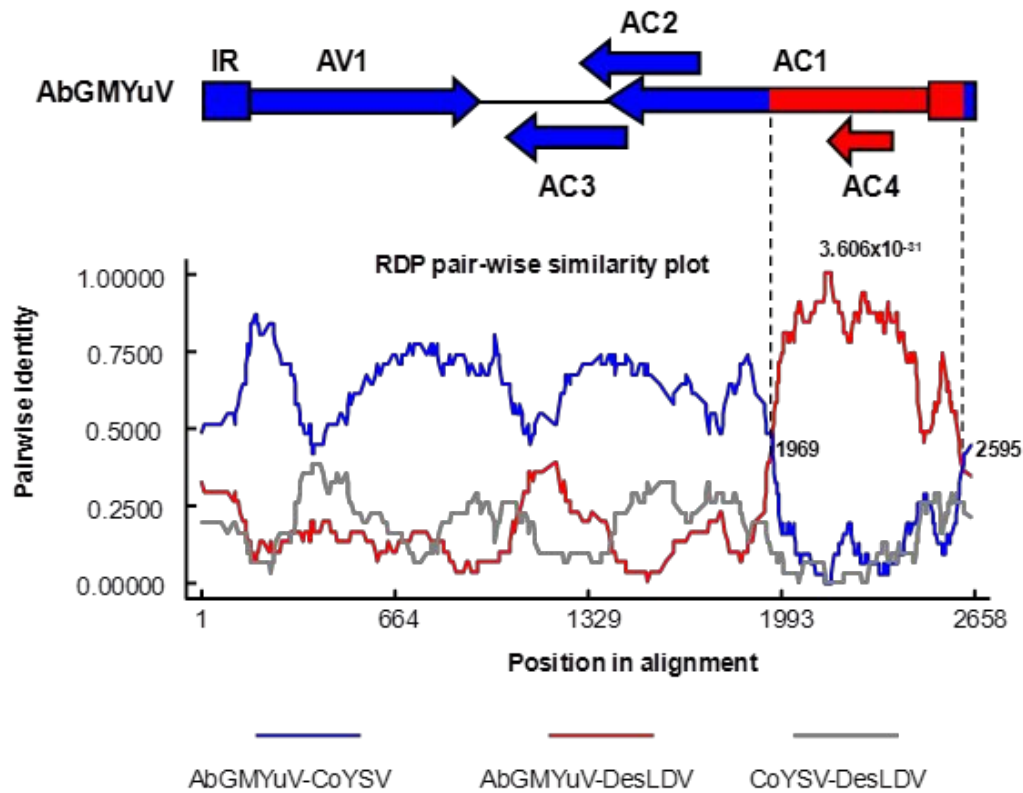


Figura 7. Evento de recombinación detectado en el DNA-A de AbGMYuV. En la parte superior de la figura se representa esquemáticamente el componente genómico A y el trazado de similitud pareada definida por el programa RDP4 se muestra en la parte inferior. La región azul indica la sección correspondiente al progenitor mayor (CoYSV), y la región en rojo al progenitor menor. Las líneas en color en el trazado de similitud. El comienzo (posición 1969) y el término (2595) de la secuencia derivada del progenitor menor se indican con líneas verticales.

Generación de clones infectivos de los begomovirus aislados

Tomando en consideración que todos los nuevos BGVs aislados en este estudio fueron aislados de plantas infectadas por dos virus, fue necesario obtener clones infectivos de los mismos para establecer sus características patogénicas y la inducción de síntomas en infecciones únicas. En consecuencia, se generaron clones de los componentes genómicos A y B completos de JacLDYuV, CapYSYuV

y ViYMV, utilizando oligonucleótidos especiales diseñados sobre sitios de restricción únicos en cada componente viral, con sus extremos 5' complementarios (iniciadores cola-con-cola) [ver Materiales y Métodos]. Una vez obtenidas y verificadas por secuenciación las clonas de cada componente genómico, se inocularon plantas de *Nicotiana benthamiana* en el estadio de 4-6 hojas por medio de biobalística. La inoculación con las clonas de JacLDYuV produjeron síntomas muy claros a los 10 dpi en las 8 plantas bombardeadas, las cuales mostraron un acortamiento muy notorio de los internodos, amarillamiento y deformación severa de las hojas emergentes (**Fig.8a y 8d**). Los síntomas no disminuyeron con el tiempo, y las plantas no desarrollaron flores aún a los 40 dpi, cuando se dió por concluído el experimento. En contraste, 6 de las 8 plantas inoculadas con las clonas de CapYSYuV desarrollaron síntomas menos severos, incluyendo el rizado leve de las hojas y un retraso moderado en el crecimiento respecto a las plantas control (**Fig. 8b y e**); los síntomas disminuyeron a partir de los 20 dpi, sugiriendo un proceso de remisión. Por último, las clonas de ViYMV produjeron síntomas en las plantas bombardeadas (6/8), que exhibieron clorosis y “arrugamiento” moderado en las hojas a los 12 dpi, y retraso en el crecimiento respecto al control, aunque también se observó remisión de síntomas a partir de los 25 dpi (**Fig. 8c y f**). En todo los casos se verificó la presencia del DNA viral en las hojas más jóvenes a los 15 dpi. Tomadas en conjunto, estas observaciones indican que los tres BGVs inoculados tienen una capacidad diferenciada para inducir síntomas en el mismo huésped (*N. benthamiana*) y confirmaron la naturaleza complementaria de los componentes genómicos aislados, lo que era particularmente importante establecer en el caso de CapYSYuV, cuya región

común exhibe una identidad en secuencia inusualmente baja entre componentes genómicos cognados, de tan solo 72%, probablemente la más baja reportada en un begomovirus bipartita; sin embargo, los elementos relevantes del origen de replicación, como los iterones y la secuencia tallo-asa, se encuentran bien conservados en este virus.



Figura 8- Síntomas inducidos por las clonas generadas de los componentes genómicos A y B de JaCLDYuV (a y d), CapYSYuV (b y e), y VYMV (c y f).

Determinantes de especificidad replicativa de los begomovirus aislados

El análisis del origen de replicación de los BGVs caracterizados en este trabajo mostró que todos ellos exhiben el arreglo de iterones característico de los miembros del clado SLCV, aunque los mismos presentan una secuencia consenso

que difiere en uno o dos nucleótidos. Por ejemplo, CapYSYuV exhibe 6 iterones con el consenso TGGAGTCC, en tanto que los componentes genómicos de JacLDYuV y AbGMYuV tienen 4-5 iterones con la secuencia TGGTGTCC. Ambos tipos de secuencias iterativas son las que predominan entre los miembros del linaje referido (Gregorio-Jorge *et al.*, 2010). En contraste, VYMV posee 5 iterones con una secuencia TGGAGACC, que difiere en el 6º nucleótido de los elementos equivalentes del resto de los virus de su linaje. De hecho, una revisión exhaustiva de los iterones de otros virus del continente americano reveló que los iterones de VYMV no se encuentran en ningún otro virus del NM descrito al presente, aunque el núcleo GGAGAC se encuentra con frecuencia en los sitios de unión de la proteína Rep de BGVs de Africa y Asia (v.gr: ACMV y CLCMV).

Identificación de un dominio distintivo de la proteína Rep de los miembros del clado del SLCV

Como se señaló previamente, los virus del linaje del SLCV se distinguen de otros BGVs por el número y arreglo peculiar de sus elementos iterativos, y por una proteína Rep que diverge significativamente de sus homólogos en otros BGVs (Arguello-Astorga *et al.*, 1994; Arguello y Ruiz, 2001; Gregorio-Jorge *et al.*, 2010). Sin embargo, algunos BGVs como *Tomato golden mottle virus* (ToGMoV), *Tomato chino La Paz virus* (ToChLPV) y *Tobacco yellow crinkle virus* (TbYCrV), que no presentan ninguna de las características mencionadas, han sido ocasionalmente incluidos en el clado del SLCV por algunos autores (Rojas A *et al.*, 2005; Hernández-Zepeda *et al.*, 2010; Bañuelos-Hernández *et al.*, 2012), lo que podría sugerir que los rasgos antes mencionados no sean, realmente, características

específicas de linaje. Para establecer si todas las especies que se agrupan en el clado del SLCV de acuerdo a la filogenia reconstruida en este estudio (que excluye a los tres BGVs antes mencionados) poseen otras características moleculares que pudiesen ser utilizadas como marcadores de relación filogenética, llevamos a cabo un análisis comparativo minucioso de las proteínas codificadas en su componente genómico A, a saber, CP, Rep, TrAP, REn and AC4 con sus homólogos en otros BGVs del Nuevo Mundo. Se alinearon las correspondientes proteínas de los miembros del primer linaje con el resto de los BGVs de América (>50 especies) y se buscaron sistemáticamente motivos de secuencia aa conservados de modo exclusivo en los BGVs de la estirpe del SLCV. El análisis reveló varios motivos de ese tipo solo en las proteínas Rep y AC4, en tanto que las proteínas TrAP, REn y CP fueron indistinguibles de las de BGVs de otros linajes secundarios. En efecto, se identificaron varios motivos diferenciales a lo largo de los primeros 160 residuos aa de Rep, pero no en otros dominios. La ampliación del análisis comparativo para incluir las proteínas homólogas de los BGVs nativos del Viejo Mundo, como Legumovirus, Corchovirus, Sweepovirus, y otros linajes de virus bipartitas y monopartitas de Africa, Madagascar, Europa, Asia y Oceanía (Bridson *et al.*, 2010) demostró que los motivos identificados en los miembros del clado SLCV no están presentes en ningún otro subgrupo conocido del género *Begomovirus* (**Fig. 9**).

En vista de estos datos, que sugieren un evento de transferencia horizontal de secuencias entre un begomovirus ancestral del NM y un geminivirus de otro género conocido (o por conocer), decidimos extender el análisis comparativo de las proteínas Rep a todos los miembros descritos de la familia *Geminiviridae*,

incluyendo algunos reportados recientemente, cuyas características genómicas son tan diferentes a los geminivirus conocidos, que se ha propuesto formar nuevos géneros con ellos (Bernardo *et al.*, 2013; Loconsole *et al.*, 2012; Poojari *et al.*, 2013). Este análisis reveló la existencia de solo dos geminivirus fuera del clado SLCV que codifican proteínas Rep con los motivos atípicos de este linaje: los curtovirus *Horseradish curly top virus* (HrCTV) y *Spinach severe curly top virus* (SpiSCTV).

	SL-C Motif 2	RCR Motif II	SL-C Motif 3
CapYSYuV*	29-ALQMLQSLHWAI	VKPLYIRVSREEHSDGF	PHLHCLLQLSGKSNIKDAGFFD-79
AbGMYuV*	29-ALQMLQTLWSVVK	PTYIRVAREEHSDGF	PHLHCLLQLSGKSNIKDARFFD-79
JacLDYuV*	29-VLELLQNLPTIVR	PTYIRVAREIHSDGN	PHLHILIQLSGKSNIKDERFFD-79
ViYMV*	29-ALQMLQSLWSVVK	PTYIRVSREEHADGF	PHLHCLIQLSGKPSIKDARFFD-79
SLCV	29-VLEMLLHLSWSVVK	PNYVRVAREEHSDGS	PHLHCLIQLSGKSNIKDAGFFD-79
ToSLCV	29-AIEMLQNLPSVVK	PTYIRVAREEHADGF	PHLHCLIQLSGKSNIKDARFFD-79
CaLCuV	29-ALQMLQTLWSVVK	PTYIRVAREEHSDGF	PHLHCLIQLSGKSNIKDARFFD-79
HrCTV	29-MLQLLQNLNWT	VVKPTYIRVAAERHADG	TPHLHCLIQLSGKSNIKDCRFFD-79
ToChLPV	28-ALDQLLSIQLP	SNK-KYIKICRELHENG	EHHLHVLMQFEGKIQITNQRQFD-77
PYMV	28-ALSQQLNLTIP	VNK-KFIKICRELHENG	EPHLVLIQFEGKYNCTNNRFLD-77
CoYV	26-ALEQIKTLSLPT	NI-VYIRVCEEKHQD	GSPhLHALVQFQKKYRCTNCRFLD-75
ICMV	28-ALSQIRNFQTP	TNP-KFIKICRELHENG	EPHLHVLIQFEGKYKCQNQRFFD-77
MYMV	28-ALEE L LALST	PV NK-KFIRVCRELHED	GEPHLHVLLQFEGKFQTKNERFFD-77
SACMV	28-ALDQLRQLQT	PTNK-LFIKICRELHENG	EPHLHALIQFEGKYNCTNQRFFD-77
TYLCV	26-ALSQ LKKLET	PTNK-KYIKVCKELHENG	EPHLHVLIQFEGKYQCKNQRFFD-75
TPCTV	29-ALDQLQRLQT	PTNK-KYIKVARELHENG	EPHLHVLIQFEGKFNCKNQRFFD-78

		SL-C Motif 4	RCR Motif III	SL-C Motif 5	SL-C Motif 6	
CapYSYuV*	96-	AKDTNAVRNYI	TKEGDYCE	SGQYKVS	GSKANKD	DVYHN
AbGMYuV*	96-	AKDTNAVKNYI	TKEGDYCE	SGQYKVP	GSKANKD	DVYHN
JacLDYuV*	96-	AKDTNAVKNYI	TKEGDYCE	SGQYKVS	ATRTNKE	DVYHN
ViYMV*	96-	AKDSNAVKNYI	TKEGDYCE	SGQYKVS	ATKSNKD	DVYHN
SLCV	96-	AKDTNAVKNYI	TKEGDYCE	SGQYKVS	GSKSNKD	DVYHN
ToSLCV	96-	AKDTNAVQNYI	TKEGDYCE	SGQYKVS	GTKSNKD	DVYHN
CaLCu	96-	AKDTNAVKNYI	TKDGDYCE	SGQYKVS	GSKANKD	DVYHN
HrCTV	96-	AKDTNAVKNYI	TKEGDYCE	SGQYKVS	GSKANKD	AVYHN
ToChLPV	94-	AKSSSDVKS	YIDKGD	TLEWGEFQ	IDGRSARGG	QQTAN
PYMV	94-	AKSSSDVKS	YVEKGD	TIEWGLFQ	IDGRSARGG	QQTVN
CoYVV	92-	ARSSDAKSYI	EKDGVFCE	WGEFKIDGR	SSRGGQQS	ANDAYAK
ICMV	94-	AKSSSDVKS	YIDKGD	TLEWGTFFQ	IDGRSARGG	QQSAN
MYMV	94-	AKSASDVKS	YMDKGD	VVDHGSFQ	VDGRSARGG	KQSANDAY
SACMV	94-	AKSSSDVKS	YLDKGD	TIQWGEFQ	IDGRSARGG	QQSAN
TYLCV	92-	AKSSTDVKTY	VEKDNFID	FGVFSQ	IDGRSARGG	QQSAN
TPCTV	95-	AKSSSDVNSY	VDKGD	TIEWGEFQ	IDARSARGG	QQTAN

Figura 9- Motivos en Rep específicos del linaje SLCV. Alineamiento de las regiones 29-79 y 96-137 de las proteínas codificadas por los nuevos virus y otros miembros del mismo linaje, el curtovirus HrCTV, BGVs del continente americano (ToChLPV y PYMV), de Asia (CoYVV, ICMV, MYMV) y de Africa (SACMV y TYLCV), así como la proteína homóloga del único topocovirus conocido (TPCTV). Los residuos conservados en todas las proteínas comparadas están sombreadas en azul, en tanto que los motivos distintivos de los miembros del clado SLCV están sombreados en amarillo. Un motivo de 18 aa característico de la proteína de los begomovirus típicos se encierra en el rectángulo punteado.

Para examinar desde una perspectiva más completa las relaciones entre las proteínas Rep de los parientes de SLCV con las proteínas homólogas de otros geminivirus, se generó un árbol filogenético basado en el alineamiento del dominio amino-terminal de Rep (segmento 1-160) de un gran número de virus representativos de diferentes géneros o linajes geográficos. El dendrograma resultante (**Fig. 10**) confirmó la estrecha relación entre las proteínas de HrCTV y SpiSCTV con las moléculas homólogas de los miembros del clado SLCV. Un dato de sumo interés surgido del análisis filogenético es la observación de que las

proteínas Rep de los curtovirus antes mencionados, y la de JacYMYuV, uno de los nuevos BGVs caracterizado en este estudio, son basales a la rama del SLCV, sugiriendo de ese modo una relación cercana con el dominio correspondiente del ancestro común de los virus del clado SLCV. Las proteínas de éstos últimos virus quedaron claramente separada de la rama principal de Reps de los BGVs del Nuevo y el Viejo Mundo, dentro de la cual también quedaron incluidas las proteínas homólogas de virus pertenecientes a tres géneros diferentes, *Topocuvirus* (v.gr: TPCTV), *Turncurtovirus* (v.gr: TCTV), y *Curtovirus* (v.gr: BMCTV and BSTCV). La conclusión de este estudio comparativo fue que el dominio 1-160 de la proteína Rep codificada por los miembros del clado del SLCV no se derivó de un begomovirus, sino de un geminivirus perteneciente a un linaje diferente. Puesto que existen dos curtovirus (HrCTV y SpiSCTV) que codifican una Rep con los motivos distintivos de la proteína de los BGVs del linaje SLCV, la cuestión acerca del presunto ancestro no-begomoviral de esos virus parece tener una respuesta obvia; sin embargo, un análisis minucioso mostró que la solución al enigma es más compleja de lo que aparenta, como se discutirá más adelante.

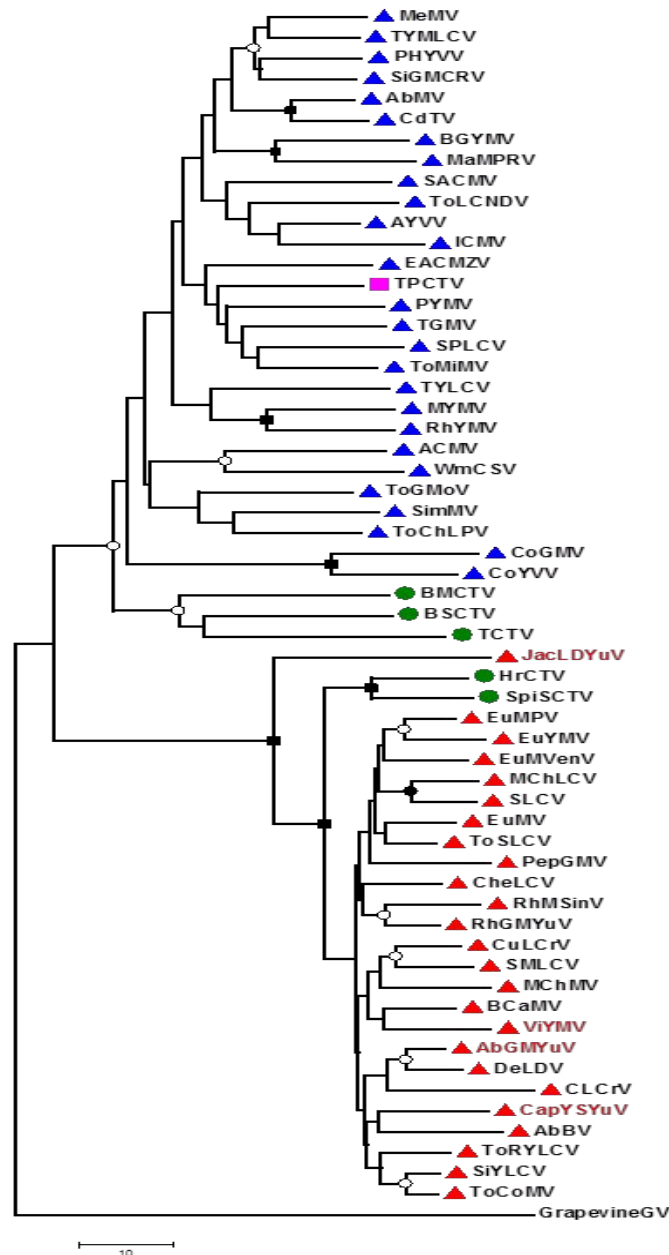


Figura 10. Filogenia derivada del dominio N-terminal de Rep. El dominio 1-160 de las proteínas Rep de una variedad de begomovirus, curtovirus, turncurtovirus y topocuvirus, fue utilizado para establecer relaciones filogenéticas entre los mismos. Los triángulos de color rojo denotan a los miembros del clado SLCV, los triángulos azules a begomovirus de otros linajes, los círculos verdes a los curtovirus y al único turncurtovirus conocido (TCTV), y el cuadrado en rosa al único miembro del género *Topocuvirus*, TPCTV. El árbol fue enraizado con la secuencia del dominio equivalente de “Grapevine geminivirus”, un miembro muy divergente de esta familia viral que no ha sido oficialmente clasificado en ningún género.

Análisis de la proteína AC4 de los miembros del clado del SLCV

Dado que el gen *AC4* de los begomovirus se encuentra traslapado en su totalidad con el segmento 5' del gen *AC1*, que codifica a los primeros 160 aminoácidos de Rep, se puede anticipar que la proteína AC4 de los BGVs de la estirpe del SLCV será muy diferente a las proteínas equivalentes de los begomovirus de otros linajes. En efecto, el alineamiento de esas proteínas reveló una clara divergencia en secuencia entre las codificadas por los miembros del clado SLCV y las moléculas homólogas de otros begomovirus, con una identidad en secuencia de aminoácidos inferior al 28%. Además de la baja similitud en secuencia, las proteínas AC4 de SLCV y sus parientes fueron consistentemente más grandes, con una longitud promedio de 120 aminoácidos, que contrasta con los ~85 residuos de los que constan las proteínas AC4/C4 de los begomovirus de otros linajes. El alineamiento también reveló la existencia de motivos conservados específicos del linaje SLCV, que también se observan en las proteínas de los curtovirus HrCTV y SpiSCTV, como cabía esperar (**Fig. 11**).

La filogenia reconstruida a partir de las proteínas AC4/C4 de begomovirus, curtovirus, topocuvirus y turncurtovirus (**Fig. 12**) reflejó de modo cercano la filogenia derivada de las proteínas Rep truncadas (1-160). Así pues, el mayor tamaño y la secuencia divergente de las proteínas AC4 constituyen otros rasgos moleculares distintivos de los BGVs del clado del SLCV, lo que sugiere la existencia de diferencias funcionales con las proteínas homólogas de otros begomovirus.

```

CapYSYuV* 51-SPTSRLDFSTLLTPEGLPVFTQISKLPKTTPSEITS PKRVIIVNPGSIRCLGVP-106
AbGMYuV*  43-SPTSRLDFSTSLTQEGPLPFTQISRLPKTTPSRITS PKRVIIVNPDNTRCLAEQ-98
JacLDYuV*  42-NQTSRMNDFSTYLTHQGLPLSTLISKQPKTTPSRITS PKRVIIVNPDNTRCLGQQ-97
ViYMV*    43-SPASRLDFSTSLTPEGLPVFTQFKQPKIQMQSRITS PKRVIIVNPGSTKCLGQQ-98
SLCV      47-NPTSRLDFSTSLTPDGLPDTQTFRLPKTTPSRITS PKKVIIVNPGNIKCLGVQ-102
ToSLCV    43-SPTSRLDFSTSLTPEGLPIFIQTHRQPKTTPSRITS PKKVIIVNPDNTKCLGVQ-98
CaLCuV    43-SRTSRMLDFSTSLTPEGLPIFTQTFRQPKTPMPSRITS PKMVIIVNPGSTRCLGVQ-98
HrCTV     43-SPTSRTADFSTLLTPNGVPLSILMSNQPKTPMPSRITS PKKVIIVNPDNTRSLGEQ-98
ToChLPV   15-KSKSRIKDSSTWYPQPGQHISIRTFRELNQAPTSSPTSTRMEIHWNGENSRSTEDL-70
PYMV      15-NTTAQITDCSIWCPQPGQHISIRTFRELNRAPMS SRMSRKTETPSNGVCSRLTEEV-70
CoYVV     15-STDARTADCSIYPIQIHNNNTIVYR LHAPPPTPNRTSKKMEYSANGVNSRSTADH-70
ICMV      17-NTNARISDSSTWYPHPGQHISIRTFKELNPAPTSSPTSTKTAIHWNGEHSRSMEDL-72
MYMV      17-SSKRRTKGSSTWFPQPDQHITIRTFRQLKAHQMLSHTWTKTETSLTMEVSKSMADQ-72
SACMV     17-STIVPTNDSSTSYPLPGQHISIQTFRELNPLVLTSSPIWTRTETPSNGASFRSTDDL-72
TYLCV    15-NTNVRTNGSSTWYPQTGQHISIRTFRQLRAQQMSRPTWRKTETSLILEFPKSMADQ-70
TPCTV    15-NSTAKINDSSTWYPQPGQHISIRTYRELNPAPTSTPTSTRTETLSNGENSRLLDL-70

```

Figura 11. Motivos conservados en las proteínas AC4 de begomovirus del clado SLCV. Alineamiento de la región 43-98 de AC4 de virus del linaje mencionado y del curtovirus HrCTV, con la región equivalente (segmento 15 a 70) de AC4/C4 de otros begomovirus y TPCTV. Los motivos distintivos de los miembros del clado SLCV están sombreados en amarillo. Nótese la baja conservación en secuencia de las proteínas codificadas por BGVs de América, Asia y África.

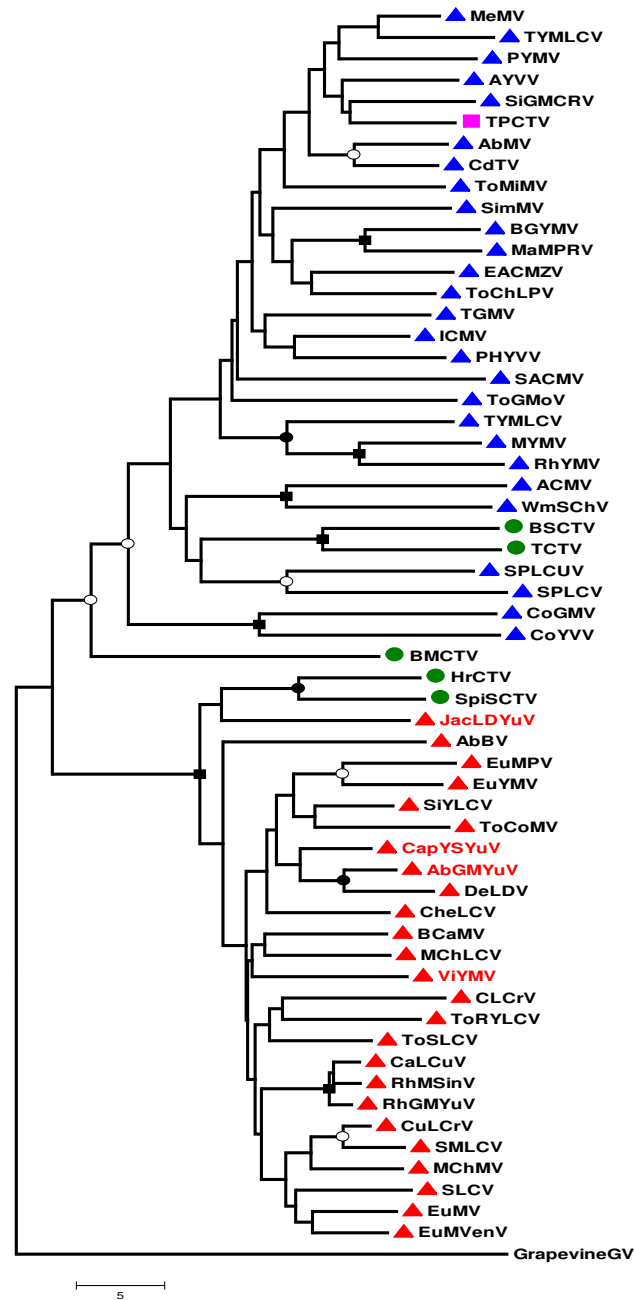


Figura 12- Filogenia de proteínas AC4/C4. Árbol filogenético que muestra las relaciones evolutivas entre las proteínas AC4 de los virus aislados en este trabajo (AbGMYuV, CapYSYuV, JaLDYuV y ViYMV) y las proteínas AC4/C4 de begomovirus selectos. Los miembros del clado SLCV se denotan con un triángulo rojo adelante del acrónimo de la especie viral. Las proteínas C4 de curtovirus (BMCTV, BSCTV, HrCTV, SpiSCTV) y el turncurtovirus TCTV se indican con un círculo verde; y la del Topocuvirus TPCTV con un cuadrado rosa.

Discusión

Existe consenso entre los expertos en paleovirología que los BGVs aparecieron en el período Crétacico, poco antes del fraccionamiento tectónico del supercontinente Gondwana, que originó los bloques continentales de Sudamérica, la Antártica, Africa, India y Australia, evento que tuvo lugar hace unos 120-110 MYA (Rybicki, 1994; Briddon *et al.*, 2010; Nawaz-ul-Rehman & Fauquet, 2009). Esa hipótesis se basa en la observación de que las filogenias basadas en la secuencia de los componentes A completos de los BGVs producen linajes continentales bien diferenciados, lo que sugiere que éstos evolucionaron en aislamiento geográfico durante extensos períodos. Los BGVs de América forman una gran rama filogenética claramente separada, y diversas líneas de evidencia indican que se derivaron de BGVs del VM, muy probablemente a partir de la estirpe de los “Corchovirus”, de India y Vietnam (Ha *et al.*, 2006; 2008). Se suele asumir que los BGVs se introdujeron al continente americano por la región de Alaska (Lefevre *et al.*, 2010; Briddon *et al.*, 2010), aunque se ha postulado una hipótesis alternativa que propone que los BGVs arribaron al Nuevo Mundo por Sudamérica, provenientes del continente Antártico (Cárdenas-Conejo, 2011; Cárdenas-Conejo *et al.*, en preparación). En cualquier caso, tuvo lugar una radiación evolutiva de los BGVs en el curso de su propagación por el continente, lo que dio origen a linajes secundarios como los clados del AbMV, de BGMV, de PHYVV, y del SLCV, entre otros (Briddon *et al.*, 2010). En este estudio logramos establecer, a través de la caracterización molecular de nuevas especies del clado SLCV, que este linaje es diferente a otras estirpes del Nuevo Mundo en rasgos moleculares relacionados

con la replicación viral, como se hizo notar 20 años atrás (Argüello-Astorga *et al.*, 1994a, b). En relación al gen AC4, totalmente traslapado con el gen *AC1/rep*, lo más notable es la gran diferencia en secuencia y tamaño de la proteína que codifican los miembros del clado del SLCV, que es ~35 residuos aa más larga que las proteínas homólogas de otros BGVs, del topocuvirus TPCTV, del turncurtovirus TCTV, y del curtovirus BCTV, los cuales codifican proteínas AC4/C4 con una longitud promedio de 85 residuos aa. Al presente, y en ausencia de estudios con mutantes AC4 de miembros del clado del SLCV, nada se puede afirmar acerca de las funciones de la proteína correspondiente, aunque llama la atención el hecho de que uno de los BGVs mejor estudiados de este clado, PepGMV, carece de un ORF AC4. Esta parece ser la única excepción entre las 30 especies virales del linaje en cuestión.

Determinantes de especificidad replicativa de los miembros del Clado SLCV

El análisis desarrollado en este y otros trabajos previos de nuestro grupo ha mostrado que los miembros del clado SLCV poseen una región intergénica que exhibe diferencias importantes con la de otros BGVs del continente americano: 1) no presentan la caja G asociada al extremo 5' del elemento tallo-asa del Ori, que es característica de los BGVs americanos, y en vez de eso exhiben una secuencia conservada GGGGYAAAA, cuya función se desconoce; 2) exhiben de 4 a 6 iterones, lo que contrasta con los tres iterones observados en la mayoría de los BGVs del NM (ver Fig.2); 3) sus iterones presentan el consenso TGGWGC (una excepción es RhYMSinV, con iterones TGGAGGA), y nunca están precedidos por el motivo AATY que comúnmente se presenta asociado a la secuencia nuclear de

los iterones de begomovirus “típicos”. Más aún, el análisis de los potenciales SPDs de las proteínas Rep codificadas por los virus del linaje SLCV mostró que aquellos correlacionan, clara y consistentemente, con diferencias en secuencia nucleotídica de los iterones cognados, como se ilustra en la **Fig. 13**.

	SPDs-R1		Motif I		Motif II	SPDs-R2					
ITERON	MPRAPGS	FRLA	AKNI	FLTYPQC	---36---	HLHCLL	QLSGKS	NIK	DARFFD	AbGMYuV	
	MPRQPNK	FRLA	SRNI	FLTYPQC	---36---	HLHILI	QLSGKS	NIK	DERFFD	JaeLDYuV	
	MPRNPSS	FRLT	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DVRFFD	BCaMV	
	MPRNPKS	FRLA	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	CaLCuV	
	MPRNPNS	FRIS	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	CuLCrV	
	MPLKPKT	FRLA	ARNI	FLTYPQC	---36---	HLHCLL	QLSGKS	NIK	DARFFD	DeLDV	
	MPRNPNS	FRLT	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	EuMPV	
	MPRNPNS	FRLA	AKSI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	EuMV-YP	
	MPRNPNS	FRLT	AKNI	FLTYPRC	---36---	HLHCLI	QLSGKS	NIK	DARFFN	EuYMV	
	MPQNPKS	FRLS	ARNI	FLTYPKC	---36---	HLHCLV	QLSGKS	NIK	DAGFFD	MeChLCuV	
	MPRS PKS	FRVA	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	NARFFD	MeChMV	
	MPREP KS	FRLA	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	NARFFD	SiGMLV	
	MPRNPQK	FRLA	AKNI	FLTYPQC	---36---	HLHCLV	QLSGKS	NIK	DVRFFD	SiYLCV	
	MPRNPNS	FRLT	ARNI	FLTYPRC	---36---	HLHCLI	QLSGKS	NIK	DAGFFD	SLCV	
	MPRNPNS	FRFT	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKT	NIK	NARFFD	SMLCV	
	MPRNPNI	FRLT	ARNI	FLTYPQC	---36---	HLHCLV	QLSGKS	NIK	DVRFFD	ToCoMV	
	MPRNPNQ	FRLA	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	ToRYLCV	
	MPRNPNL	FRLA	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	ToSLCV	
		MPRQNK S	FRL S	SKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DCRFFD	SpSCTV
	TGGTGTCC	MPRNPNS	FRLQ	ANNI	FLTYAQC	---36---	HLHCLL	QLSGKS	NIK	DAGFFD	CaYSYuV
MPRLPHS		FRLQ	ARNI	FLTYPRC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	AbMV	
MPRNPNS		FRIQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	CheLCV	
MPRNPDS		FRLQ	ARHI	FLTYPKC	---36---	HLHCLI	QLSGKS	NIK	NKRFFD	CLCrV	
MPRLPNS		FRLQ	ARNI	FLTYPRC	---36---	HLHCLV	QLSGKS	NIK	DARFFD	CoMoV	
MPRQPK E		LRVQ	SKNI	FVTPQC	---36---	HLHCLV	QLSGKP	NIK	DVRFFD	CoChSV	
MPRNPNS		FRIQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	EuMV-Jal	
MPRNPNS		FRLQ	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	EuMVenV	
MPLPKS		FRLQ	CKNI	FLTYPQC	---36---	HLHCLI	QLTGKC	NIK	DARFFD	PepGMV	
MPRNPNS		FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	RhGMYuV	
MPRQPNT		FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	TBYMV*	
		MPRQPNS	FRIQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DCRFFD	HrCTV
TGGAGACC		MPPKPSA	FRIQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKP	SIK	DARFFD	ViYMV
TGGAGGAC	MPRNPNS	FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGES	DIK	NARFFD	RhMSinV	

Figura 13. Alineamiento de la región N-terminal de proteínas Rep de los 30 miembros conocidos del clado SLCV y los dos curtovirus (secuencia en letras rojas) que poseen proteínas similares. La secuencia consenso de los iterones cognados se muestra en el lado izquierdo de la figura. Los acrónimos a la derecha identifican a la especie viral. Los cuatro residuos aminoácidos que presumiblemente están involucrados en la especificidad de unión al DNA de la proteína están sombreados en amarillo, salvo en dos casos. Los acrónimos sombreados de amarillo corresponden a los virus caracterizados en este estudio.

¿Cuál es el origen del segmento genómico distintivo de los miembros del clado SLCV?

La filogenia derivada de la comparación de los primeros 160 residuos aa de las Reps de miembros del linaje del SLCV, de otros begomovirus americanos y del VM, y los homólogos codificados por miembros de otros tres géneros de geminivirus (*Curtovirus*, *Turncurtovirus* y *Topocuvirus*) (Fig.10) reveló la existencia de dos amplias clases de geminivirus: 1) los que codifican proteínas Rep de tipo begomoviral “típico” (BGVs del NM y VM; el curtovirus BCTV y sus cepas BMCTV y BSCTV; y las especies únicas de los géneros *Turncurtovirus* y *Topocuvirus*); 2) los que codifican los miembros del clado SLCV y los curtovirus HrCTV y SpiSCTV. Estos dos grandes grupos pueden ser representados de modo abreviado, pero exacto, por los motivos conservados en la región ~125 de Rep, que en el caso de los BGVs típicos es DGRSARGGQQ, en tanto que en la estirpe del SLCV es YKVSKGRKA (**Fig. 14**).

Dado que los virus del linaje del SLCV son, sin duda, miembros del género *Begomovirus*, por su genoma bipartita, y la secuencia de sus proteínas CP, TrAP (o AC2), REn (o AC3), y el segmento carboxi-terminal (161-360) de Rep, surge la interrogante de qué geminivirus fue el que aportó, vía recombinación intermolecular, el segmento de ~700 nt que distingue a los miembros del clado del SLCV. Con base en el análisis comparativo que realizamos, en el que se identificaron dos curtovirus con un segmento homólogo al de los parientes de SLCV, la hipótesis más simple es que un ancestro de esos dos curtovirus, HrCTV y SpiSCTV, fue el “donador” del segmento en cuestión. Sin embargo, un examen más profundo de la cuestión demostró que esa suposición no es correcta, por dos

razones sólidas: 1) La semejanza de la proteína Rep de HrCTV y SpiSCTV con sus homólogos del clado SLCV no se limita a los primeros 160 residuos aa, sino que se extiende a lo largo de toda la proteína. En efecto, el dominio 161-346 de la proteína de SpiSCTV que tiene una identidad en secuencia del 75% con la Rep de HrCTV, exhibe una identidad similar o incluso superior (80%-73%) con el dominio homólogo de la Rep de BGVs del Nuevo Mundo, y significativamente menor (66-60%) con otros curtovirus (todas las cepas de BCTV) y los BGVs del Viejo Mundo.

2) El ORF “C2” traslapado con la parte final del gen AC1 de HrCTV y SpiSCTV codifica una proteína no conservada, que exhibe tan sólo un 28% de identidad en secuencia de aa, y no presenta similitud significativa con ninguna otra proteína, viral o no viral, en las bases de datos públicas. No obstante esas observaciones, el mencionado ORF “C2” parece haberse derivado de un gen (A)C2 begomoviral, ya que tiene la misma extensión (134 codones) que el gen de los BGVs, y además 9-8 de los primeros 12 residuos de la proteína “C2” de SpiSCTV son idénticos a los 12 aa iniciales de la proteína (A)C2 de algunos BGVs de Africa (v.gr: EACMV), China (v.gr: AYVCNV) Pakistan (v.gr: CLCAV) y Vietnam (v.gr: PrLCV), lo que contrasta marcadamente con la baja identidad en secuencia que exhibe respecto a la proteína homóloga de HrCTV (**Fig. 15**). Esas observaciones son compatibles con la hipótesis propuesta por Stenger et al (1996) de que HrCTV, y probablemente también SpiSCTV (Zepeda-Hernández y Brown, 2010) evolucionaron por un evento de recombinación entre un curtovirus y un begomovirus del linaje del SLCV, que habría sido el “donador” del gen *AC1/rep*. En cambio, esos datos no apoyan en modo alguno la suposición de que un

curtovirus parecido a HrCTV y SpiSCTV haya sido el donador no begomoviral del segmento genómico distintivo de los virus del clado del SLCV.

Estas y otras evidencias muestran que no se ha descrito, hasta el momento, un geminivirus que pueda considerarse representativo del “progenitor menor” del ancestro común de los virus del clado SLCV, y lo único que sabemos del primero es lo que podemos inferir de las características comunes del origen de replicación, el ORF AC4 y los primeros 160 codones del gen *AC1/rep*, tal y como lo encontramos en sus descendientes modernos.

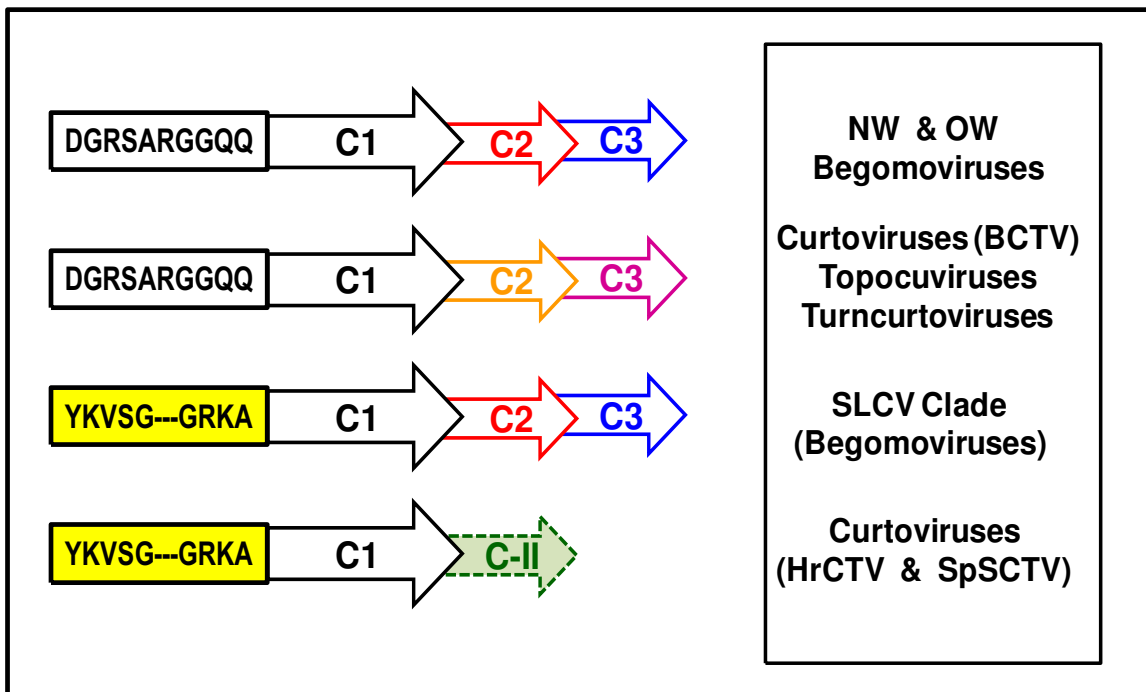


Figura 14. Dos clases de proteínas Rep caracterizan a los geminivirus de 4 géneros.
(Explicación en el texto)

<p>AC2 vprotein [East African cassava mosaic virus] Sequence ID: emb CBA13492.1 </p> <p>Query 1 MPYSFPSVNHCT 12 MP S PS NHCT</p> <p>Sbjct 1 MPSSSPSKNHCT 12</p>	<p>AC2 protein [Cotton leaf curl Alabad virus] Pakist Sequence ID: ref NP_795360.1 </p> <p>Query 1 MPYSFPSVNHCT 12 M S PS NHCT</p> <p>Sbjct 1 MQSSSPSRNHCT 12</p>
<p>AC2 protein [Tomato yellow leaf curl Yunnan virus] Sequence ID: gb AGM20539.1 </p> <p>Query 1 MPYSFPSVNHCT 12 M S PS NHCT</p> <p>Sbjct 1 MQNSSPSRNHCT 12</p>	<p>AC2 [Ageratum yellow vein China virus] Sequence ID: gb AHH24773.1 </p> <p>Query 1 MPYSFPSVNHCT 12 M S PS NHCT</p> <p>Sbjct 1 MQNSSPSRNHCT 12</p>
<p>AC2 [Premna leaf curl virus] Vietnam Sequence ID: gb AFM44664.1 </p> <p>Query 1 MPYSFPSVNHCT 12 M S PS NHCT</p> <p>Sbjct 1 MQNSSPSRNHCT 12</p>	<p>C2 [Horseradish curly top virus] Sequence ID: gb U49907 </p> <p>Query 1 MPYSFPSVNHCT 12 MP H</p> <p>Sbjct 1 MPFLYSWILHYS 12</p>

Figura 15. El ORF C2 del curtovirus SpiSCTV proviene evolutivamente del gen AC2 de un begomovirus. Alineamiento de los primeros 12 residuos aa de la proteína codificada por el ORF C2 de *Spinach severe curly top virus*, con proteínas homólogas de algunos begomovirus y del curtovirus HrCTV.

En relación a las características del “progenitor mayor” del ancestro recombinante, se puede inferir de la secuencia del JacDLYuV (el miembro del clado del SLCV presumiblemente más cercano al ancestro común) que fue un begomovirus cercano al tronco inicial que colonizó el Nuevo Mundo, pues la secuencia del componente genómico A de ese virus (excluyendo el segmento genómico atípico) presenta una similitud relativamente constante (~77%) con BGVs de diferentes estirpes americanas [datos no mostrados]. Así pues, si tomamos como punto de partida la hipótesis predominante de que los begomovirus arribaron al continente americano a través de Alaska, procedentes de Asia, tendríamos que asumir que las secuencias no begomovirales de los miembros del linaje del SLCV provienen

de una clase de geminivirus hasta ahora desconocida, pero con toda probabilidad nativa del continente americano, ya que solo se han encontrado sus secuencias en virus del Nuevo Mundo. Ahora bien, la divergencia existente en el segmento “atípico” del ADN-A de los virus del clado del SLCV indica que éste linaje es antiguo, tanto o más que cualquier otro linaje americano, si se considera la divergencia en el segmento genómico análogo. Si los estimados paleoviroológicos de la antigüedad de los BGVs en América, que oscilan entre 30 y 4 millones de años (Lefebvre et al., 2010; Murad et al., 2004) son correctos, la estirpe del SLCV debió evolucionar antes de la formación del Istmo de Panamá (hace ~3 MYA), que dio lugar al intercambio masivo de fauna y flora entre Norteamérica y Sudamérica. Esto conduce a una nueva cuestión: el geminivirus que transfirió horizontalmente el segmento Ori-Rep160 al BGV ancestral de la estirpe del SLCV ¿era nativo del bloque continental de Norteamérica o de Sudamérica? Por razones que son ampliamente discutidas por Cárdenas-Conejo (2012), y que no reproducimos aquí, consideramos más plausible la hipótesis de que el mencionado ancestro haya sido nativo de Sudamérica, y por ende, postulamos que el linaje del SLCV se originó ahí y migró a Norteamérica solo después de la formación de un puente terrestre entre los dos subcontinentes. Esta hipótesis es contraria a la sostenida por la mayoría de los especialistas en evolución de los geminivirus (Briddon et al., 2010; Lefebvre et al., 2011; Rojas et al., 2005; Rybicki, 1994). Los corolarios evolutivos de nuestra hipótesis podrán ser o no confirmados por la investigación futura, pero en el estado actual del conocimiento parecen integrar de manera más coherente las evidencias reunidas en este y otros estudios sobre la evolución de los begomovirus en el Nuevo Mundo.

REFERENCIAS

- Adams, M.J., King, A.M.Q. & Carstens, E.B. (2013). Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2013). *Arch Virol* 158, 2023-2030.
- Argüello-Astorga, G.R., Guevara-González, R.G., Herrera-Estrella, L.R. & Rivera Bustamante, R.F. (1994a). Geminivirus replication have a group-specific organization of iterative elements: A model for replication. *Virology* 203, 90-100.
- Argüello-Astorga, G.R., Herrera-Estrella, L.R. & Rivera Bustamante, R.F. (1994b). Theoretical and experimental definition of geminivirus origin of replication. *Plant Molecular Biology*, 26: 553-556.
- Argüello-Astorga, G.R. & Ruiz-Medrano, R. (2001). An iteron-related domain is associated to motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol* 146, 1465-1485.
- Argüello-Astorga, G.R., Ochoa-López, L., Kong, Ling-Jie., Orozco, M., Settlage, S.B., Hanley-Bowdoin, L. (2004). A novel motif in geminivirus replication proteins interacts with the plant retinoblastoma-related protein. *J Virol* 78, 4817-4826.
- Argüello-Astorga G. et al. (2007) High-frequency reversion of geminivirus replication protein mutants during infection. *Journal of Virology* (20) 11005-11015
- Bañuelos Hernández, Bernardo (2012) Identificación de los determinantes de especificidad de la proteína iniciadora de la replicación de geminivirus por mutagénesis dirigida. Tesis Doctoral.
- Böttcher, B., Unseld, S., Ceulemans, H., B-Russell, R., Jeske, H. (2004). Geminiate Structures of African Cassava Mosaic Virus Geminiate Structures of African Cassava Mosaic Virus. *J Virol* 78, 6709–6714.
- Brown JK et al. (2002) *Cucurbit leaf curl virus*, a new emergent begomovirus species able to form viable reassortants with related viruses in the *Squash leaf curl virus* cluster. *Phytopathology* 92:734–742
- Briddon, R.W., Patil, B.L., Bagewadi, B., Nawaz-ul-Rehman, M.S., Fauquet, C.M. (2010). Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evol Biol* 10, 97.

- Campos-Olivas, R., Louis, J.M., Clerot, D., Gronenborn, B. & Gronenborn, A.M. (2002). The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. *Proc. Natl. Acad. Sci. USA* 99, 10310-10315.
- Chandler M. et al. (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nat Rev Microbiol* 11:525-538.
- Dellaporta, S.L., Wood, J. & Hicks, J.B. (1983) A plant DNA miniprep: version II. *Plant Mol Biol Rep* 1, 19-21.
- Eagle P., Hanley-Bowdoin L (1997) Cis elements that contribute to geminivirus transcriptional regulation and the efficiency of DNA replication. *J. Virol.* 71: 6947-6955
- Fauquet, C.M., Briddon, R.W., Brown, J.K., Moriones, E., Stanley, J., Zerbini, M., & Zhou, X. (2008). Geminivirus strain demarcation and nomenclature. *Arch Virol* 153, 783–821.
- Fondong VN (2013) Geminivirus protein structure and function. *Mol Plant Pathol.* 14: 635-49
- Fontes, E.P., Eagle, P.A., Sipe, P.S., Luckow, V.A. & Hanley-Bowdoin, L. (1994). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *J Biol Chem* 269, 8459-8465.
- Fontes EPB et al. (1994). Geminivirus replication origins have a modular organization. *Plant Cell* 6: 405–416
- Gladfelter HJ et al. (1997) Two domains of the AL1 protein mediate geminivirus origin recognition. *Virology* 239: 186–197
- Gregorio-Jorge, J., Bernal- Alcocer, A., Bañuelos-Hernández, B., Alpuche-Solís, A.G., Hernández-Zepeda, C., Moreno-Valenzuela, O., Frías-Treviño, G., Argüello-Astorga, G.R. (2010). Analysis of a new strain of *Euphorbia mosaic virus* with distinct replication specificity unveils a lineage of begomoviruses with short Rep sequences in the DNA-B intergenic region. *Virology* 7, 275.
- Gutierrez, C. (2000). DNA replication and cell cycle in plants: learning from geminiviruses. *EMBO J.* 19, 792–799.
- Gutiérrez C (2002) Strategies for geminivirus DNA replication and cell cycle interference. *Phys.Mol. Plant Pathol.* 60: 219-230

- Ha, C., Coombs, S., Revill, P., Harding, R., Vu, M., Dale, J. (2006). *Corchorus* yellow vein virus, a New World geminivirus from Old World. *J Gen Virol* 87, 997-1003.
- Ha, C., Coombs, S., Revill, P., Harding, R., Vu, M., Dale, J. (2008). Molecular characterization of begomoviruses and DNA satellites from Vietnam; additional evidence that the New World geminivirus were present in the Old World prior to continental separation. *J Gen Virol* 89, 312-326.
- Hanley-Bowdoin, L., R. Bejarano, E., Robertson, D. & Mansoor, S. (2013). Geminiviruses: masters at redirecting and reprogramming plant processes. *Nature* 11, 777-788.
- Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J. & Gronenborn, B. (1995). Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nucleic Acids Res* 23, 910–916.
- Ilyina TV, Koonin EV (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res* 20:3279–3285.
- Jeske, H. (2009). Geminiviruses. *Curr Top Microbiol-Immunol* 331,185-226.
- Jupin I et al. (1995) DNA replication specificity of TYLCV geminivirus is mediated by the amino-terminal 116 amino acids of the Rep protein. *FEBS Lett* 262: 116–120
- Kong, Ling-Jie., Orozco, B.M. Roe, J.L., Nagar, S., Ou, S., Feiler, H.S., Durfee, T., Miller, A.B., Grussem, W., Robertson, D., Hanley-Bowdoin, L. (2000). A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *EMBO J.* 19, 3485-3495.
- Laufs J et al. (1995) In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc Natl Acad Sci USA* 92:3879–3883
- Lefevre, P., Harkins, G.W., Lett, Jean-Michel, Briddon, R.W., Chase, M.W., Moury, B., Martin, D.P. (2011). Evolutionary time-scale of the begomoviruses: evidence from integrated sequences in the *Nicotiana* genome. *PLoS ONE*, 6(5): e19193.
- Loconsole G, Saldarelli P, Doddapaneni H, Savino V, Martelli GP, Saponari M. (2012) A leafhopper-transmissible DNA virus with novel evolutionary lineage in the family geminiviridae implicated in grapevine redleaf disease by next-generation sequencing. *Virology* 432:162-72.

- Londoño, A., Riego-Ruiz, L. & Argüello-Astorga, G.R. (2010). DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. *Arch Virol* 155, 1033–1046.
- Márquez-Martín B, Maeso D, Martínez-Ayala A, Bernal R, Teresa Federici M, Vincelli P, Navas-Castillo J, Moriones E. (2012) Diverse population of a new bipartite begomovirus infecting tomato crops in Uruguay. *Arch Virol.* 87:5397-413
- Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26, 2462-2463.
- Mauricio-Castillo, J.R. (2006). Métodos moleculares que potencian el descubrimiento de nuevas especies de begomovirus y la detección de infecciones mixtas. Tesis de Maestría. Instituto Potosino de Investigación Científica y Tecnológica, San Luis Potosí, SLP.
- Melgarejo TA, Kon T, Rojas MR, Paz-Carrasco L, Zerbini FM, Gilbertson RL. (2013) 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.* 87:5397-413.
- Murad L, Bielawski JP, Matyasek R, Kovarik A, Nichols RA, Leitch AR, Lichtenstein CP.(2004) The origin and evolution of geminivirus-related DNA sequences in *Nicotiana*. *Heredity* 92:352-8.
- Nash, T.E., Dallas, M.B., Reyes, M.I., Buhrman, G.K., Ascencio-Ibañez, J.T. & Hanley-Bowdoin, L. (2011). Functional analysis of a novel motif conserved across geminivirus Rep proteins. *J. Virol.* 85, 1182–1192.
- Nawaz-ul-Rehman MS, Fauquet CM.(2009) Evolution of geminiviruses and their satellites. *FEBS Lett.* 583(12):1825-32.
- Orozco, B.M., Gladfelter, H.J., Settlage, S.B., Eagle, P.A., Gentry, R.N. & Hanley-Bowdoin, L. (1998). Multiple cis elements contribute to geminivirus origin function. *Virology* 242, 346-356.
- Orozco, B.M., Miller, A.B., Settlage, S.B. & Hanley-Bowdoin, L. (1997). Functional domains of a geminivirus replication protein. *J. Biol. Chem.* 272, 9840– 9846.
- Poojari S, Alabi OJ, Fofanov VY, Naidu RA. (2013) A leafhopper-transmissible DNA virus with novel evolutionary lineage in the family geminiviridae implicated in grapevine redleaf disease by next-generation sequencing. *PLoS One.* 8(6):e64194.

- Prangishvili D, Forterre P, Garrett RA.(2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol.*;4:837-48.
- Rojas MR, Hagen C, Lucas WJ, Gilbertson RL.(2005) Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. *Annu Rev Phytopathol.* 43:361-94.
- Rybicki, E.P. (1994). A phylogenetic and evolutionary justification for three genera of *Geminiviridae*. *Arch Virol* 139, 49–77.
- Sambrook, J. & Russell. (2001). *Molecular cloning: a laboratory manual*. 3rd ed. Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY.
- Sanderfoot, A.A. & Lazarowitz, S.G. (1996). Getting it together in plant virus movement: cooperative interactions between bipartite geminivirus movement proteins. *Trends Cell Biol* 9, 353-8.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731–2739.
- Varsani, A., Martín, D.P., Castillo-Navas, J., Moriones, E., Hernández-Zepeda, C., Idris, A., Zervini, F.M. & Brown, J.K. (2014). Revisiting the classification of curtovirus based on genome-wide pairwise identity. *Arch Virol* 159, 1873-1882.
- Zhang, W., Olson, N.H., Baker, T.S., Faulkner, L., Agbandje-McKenna, M., Boulton, M.I., Davies, J.W. & McKenna, R. (2001). Structure of the Maize streak virus geminate particle. *Virology* 279, 471–477

ANEXOS

Anexo 1

Artículo publicado en *Archives of Virology*

J. A. Mauricio-Castillo[&], **S. I. Torres-Herrera[&]**, Y. Cárdenas-Conejo, G. Pastor-Palacios J. Méndez-Lozano, G. R. Argüello-Astorga^{*}. **A novel begomovirus isolated from sida contains putative *cis* and *trans*-acting replication specificity determinants that have evolved independently in several geographical lineages.** *Archives of Virology* (2014) **159**: 2283–2294.

(DOI: [10.1007/s00705-014-2073-8](https://doi.org/10.1007/s00705-014-2073-8))

& J. A. Mauricio-Castillo and **S. I. Torres-Herrera contributed equally to this work.**

Abstract

A novel begomovirus isolated from a *Sida rhombifolia* plant collected in Sinaloa, Mexico, was characterized. The genomic components of Sida mosaic Sinaloa virus (SiMSinV) shared highest sequence identity with DNA-A and DNA-B components of chino del tomate virus (CdTV), suggesting a vertical evolutionary relationship between these viruses. However, recombination analysis indicated that a short segment of SiMSinV DNA-A encompassing the plus-strand replication origin and the 5' proximal 43 codons of the Rep gene was derived from tomato mottle Taino virus (ToMoTV). Accordingly, the putative *cis*- and *trans*-acting replication specificity determinants of SiMSinV were identical to those of ToMoTV but differed from those of CdTV. Modeling of the SiMSinV and CdTV Rep proteins revealed significant differences in the region comprising the small $\beta 1/\beta 5$ sheet element, where five putative DNA-binding specificity determinants (SPDs) of Rep (i.e., amino acid residues 5, 8, 10, 69 and 71) were previously identified. Computer assisted searches of public databases led to identification of 33 begomoviruses from three continents encoding proteins with SPDs identical to those of the Rep encoded by SiMSinV. Sequence analysis of the replication origins demonstrated that all 33 begomoviruses harbor potential Rep binding sites identical to those of SiMSinV. These data support the hypothesis that the Rep $\beta 1/\beta 5$ sheet region determines specificity of this protein for DNA replication origin sequences.

Una nueva especie de begomovirus aislada de *Sida rhombifolia* posee determinantes de especificidad en *cis* y en *trans* que han evolucionado independientemente en varios linajes geográficos.

RESUMEN

En este estudio se caracterizó molecularmente a una nueva especie begomoviral aislada de una planta de *Sida rhombifolia*, colectada en Sinaloa, Mexico. Los componentes genómicos A y B de *Sida mosaic Sinaloa virus* (SiMSinV) exhibieron la mayor identidad en secuencia con los componente homólogos de *Chino del tomate virus* (CdTV), lo que sugiere una relación evolutiva directa entre estos dos virus de Sinaloa. Sin embargo, el análisis de recombinación mostró que un segmento corto del DNA-A de SiMSinV, que comprende el origen de replicación viral y los primeros 43 codones del gen que codifica a la proteína iniciadora de la replicación (Rep) se derivó, aparentemente, de un virus diferente, parecido a *Tomato mottle Taino virus* (ToMoTV), de Cuba. En correspondencia con esa observación, los presuntos determinantes de especificidad replicativa en *cis* y en *trans* de SiMSinV son idénticos a los de ToMoTV, pero diferentes a los de CdTV. El modelado *in silico* del dominio de unión al DNA de las proteínas Rep de SiMSinV y CdTV, respectivamente, reveló diferencias muy significativas en la región que incluye a la mini-hoja beta $\beta 1/\beta 5$, donde cinco potenciales determinantes de especificidad (DEs) de unión al DNA (los residuos 5, 8, 10, 69 y 71) fueron identificados previamente por métodos teóricos. La búsqueda por métodos bioinformáticos en las bases de datos públicas, condujo a la identificación de 33 begomovirus oriundos de varios continentes (América, Africa y Asia) que codifican proteínas Rep con potenciales DEs idénticos a los de la proteína Rep de SiMSinV. El análisis de la región del origen de replicación de esas 33 especies virales reveló que todas ellas poseen sitios de unión para Rep (“iterones”) con la misma secuencia nucleotídica que SiMSinV. Estos datos apoyan fuertemente la hipótesis de que la región beta1/beta5 de Rep determina las propiedades de unión específica al DNA de esa importante proteína viral.

Anexo 2

(Manuscrito en preparación)

An ancient begomovirus lineage native to the Americas displays an atypical genomic module of unknown ancestry: clues from novel weed-infecting viruses.

Sandra Iliana Torres-Herrera¹, Yair Cardenas-Conejo¹, Lina Riego-Ruiz¹, Oscar Moreno-Valenzuela², Samuel Lara-González¹, Guillermo Pastor Palacios¹, Salvador Ambriz-Granados¹, Yereni Minero-García², Armando Mauricio-Castillo³, Gerardo R. Argüello-Astorga^{1&}

¹Instituto Potosino de Investigación Científica y Tecnológica, A.C., Camino a la Presa San José 2055, 78216 San Luís Potosí, SLP, México.

²Centro de Investigación Científica de Yucatán, A.C., Mérida, Yucatán, México

³Unidad Académica de Agronomía, Universidad Autónoma de Zacatecas, Zac. 98000, México.

& Corresponding author

Abstract

The begomoviruses (BGVs) are highly diversified plant pathogens that evolved in the Old World and subsequently spread into the Americas giving rise to several lines of descent. One of these lineages, named after the *Squash leaf curl virus* (SLCV), includes a number of viral species exhibiting a unique arrangement of *cis*-acting elements in the replication origin (*Ori*) region. To get some clues about the evolution of the SLCV clade a search for additional lineage members in a variety of wild plants was undertaken, and as a result four novel viruses were characterized, including one that was basal to the group. Comparative analyses to find out molecular features differentiating the SLCV clade from other BGV lineages revealed that a ~700 bp long genomic segment encompassing the *Ori* region, the entire *AC4* gene, and the first 160 codons of the *AC1 (rep)* gene, is highly divergent from the analogous genome segment of other BGVs. Moreover, homology structural modeling of the replication protein (Rep) encoded by one of the new viruses revealed that its putative specificity determinants are located at positions slightly differing from those of its homologous in other BGV lineages. The expansion of the comparative analysis to include all Rep proteins of geminiviruses showed that the divergent domain characteristic of the SLCV lineage is only present in two members of the genus *Curtovirus*, that acquired it by horizontal DNA transfer. Collectively, the data assembled in this study indicate that the common ancestor of the SLCV clade viruses was a recombinant from both an early New World BGV and a geminivirus from a hitherto unknown lineage indigenous to the American continent.

Introduction

Geminiviruses are pathogens which cause serious diseases to food and fiber crops in all world regions with warm and temperate climates, and represent a major threat to the food security of developing countries (Rey et al., 2012; Rybicki and Pietersen, 1999). These viruses possess circular, single strand DNA (ssDNA) genomes encapsidated into virions with a twinned icosahedral morphology (Jeske, 2009). The *Geminiviridae* constitute the most diversified viral family known at present in both eukaryotic and prokaryotic systems, with over 300 species described (<http://www.ictvonline.org/virusTaxonomy.asp>). The family is taxonomically divided into seven genera (*Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus*, and *Turncurtovirus*) on the basis of genome arrangement, insect vector and host range (Adams et al 2013; Varsani et al., 2014). The genus *Begomovirus*, that encompasses all dicot-infecting geminiviruses transmitted by the cosmopolitan whitefly *Bemisia tabaci*, is by far the most diverse and ecologically successful geminivirus subgroup (Nawaz-ul-Rehman and Fauquet, 2009). Two major classes of begomoviruses (BGVs) are recognized on the basis of their genome arrangement and geographical distribution: the Old World (OW: Africa, Europe, Asia and Oceania) and the New World (NW: the Americas) BGVs (Briddon et al, 2010). The latter viruses typically possess a genome composed by two distinct circular ssDNA molecules of near 2.6 Kb, whereas the OW BGVs include both bipartite and monopartite viral species. The genomic component A (DNA-A) of the NW BGVs contains five open reading frames (ORFs), one in the virion sense (*AVI* or *CP*) encoding the capsid protein, and four in the complementary sense (*AC1* or *Rep*; *AC2* or *TrAP*; *AC3* or *REn*; and *AC4*) that encode proteins involved in virus DNA replication, temporal and fine-tuned regulation of virus

gene expression, and suppression of host antiviral responses (Fondong 2013; Hanley-Bowdoin et al., 2013). The genomic component B (DNA-B) contains only two ORFs, *BVI* and *BCI*, encoding proteins involved in the intracellular and intercellular movement of viral DNA (Rojas et al., 2005; Sanderfot and Lazarowitz, 1995). The genomic components of bipartite BGVs share a ~200-bp sequence, termed “common region” (CR), which contains the virus replication origin (*Ori*) encompassing several *cis*-acting elements closely associated to a potential ~30-nt stem-loop element with an invariant nonanucleotide (5'-TAATATTAC-3') where the virus-encoded replication initiator (Rep) introduces a strand-specific endonucleolytic cut to initiate the DNA replication by a rolling-circle mechanism (Heyraud-Nitschke et al., 1995; Lauffs et al., 1995; Singh et al., 2008).

The American BGVs are distinguished from the OW BGVs by several molecular features, including the absence of an AV2 (“pre-coat”) gene (Ha et al., 2008; Rybicki et al., 1994), a different number and arrangement of high-affinity Rep-binding sites (iterons) in the *Ori* region (Arguello-Astorga et al., 1994; Arguello-Astorga and Ruiz-Medrano, 2001), a PWRsMaGT motif in the N-terminal domain of CP (Ha et al., 2006, 2008; Harrison et al., 2002), and a conserved AVRFATDK motif in the C-terminal domain of AC3 that are absent in the OW BGVs (Mauricio-Castillo et al., 2014). Notwithstanding those differences several lines of evidence indicate that the NW BGVs evolved from OW BGVs ancestors (Bridson et al., 2010; Nawaz-ul-Rehman and Fauquet, 2009; Rybicki, 1994). A theoretical study entailing both phylogenetic reconstructions and molecular clock-based methods of comparative genome analysis, led to the hypothesis that the NW BGVs evolved from OW ancestors that arrived to the American continent 30-35 MYA, when a land bridge between North America and Eurasia, and a warm climate favoring the existence of a continuous belt of vegetation between those continents, plausibly existed (Lefevre et al., 2011). According

to this scenario, the ancestral Eurasian BGVs would have spread from Alaska into the continent engendering several secondary lineages throughout North America and, after the emergence of the Panama isthmus (3 MYA), also all through South America. One of those lineages, named after the *Squash leaf curl virus*, encompasses a number of viral species which are distributed from Southern USA to Uruguay and Argentina (Bridson et al., 2010; Brown et al., 2002; Gregorio-Jorge et al., 2010; Marquez-Martin et al., 2012). Members of the SLCV clade exhibit two molecular features distinguishing them from other NW BGVs: a) the iterons are more numerous and display a unique arrangement in the *Ori*, with two extra iterons of opposite polarity flanking the stem-loop element (Arguello-Astorga 1994a,b); b) the N-terminal domain of its Rep protein displays low sequence identity with other BGV Reps (Arguello-Astorga and Ruiz-Medrano, 2001; Gregorio-Jorge et al., 2010). The evolutionary and functional significance (if any) of those distinctive features has not been addressed until now, in spite of their importance for a better comprehension of the evolutionary history of BGVs in the American continent. In this work we report the molecular characterization of four novel viruses belonging to the SLCV clade (SL-C, for simplicity) whose genomic sequences, together with all BGVs genomes available in public databases on 30 November 2014, were examined to define with precision the distinctive molecular features of the members of this American lineage. This analysis led to the intriguing conclusion that the SL-C common ancestor probably evolved from a recombination event between an ancestral NW BGV and a geminivirus from a hitherto unknown lineage, whose *Ori* region and N-terminal Rep protein sequences have been only found so far in viruses natives to the Americas.

Results

Isolation and identification of novel begomoviruses from Mexico

Leaf samples of weeds exhibiting symptoms of virosis were collected during the summer of 2007 in several localities of the states of Yucatan and Morelos, as a part of a wide-ranging study to explore the diversity of BGVs in Mexico. The presence of BGVs was tested by PCR using both “universal” primers and SLCV lineage-specific primers (see Methods) that were designed to improve the detection of SL-C members in mixed infections (Mauricio-Castillo et al. ; Gregorio-Jorge et al., 2010). A number of the examined field samples contained two distinct BGVs, and four of these doubly-infected plants (samples YP#122, YP#446, YP#534 and Mor# 991) produced amplicons with the SL-C specific primers that were different to BGV sequences in public databases. The full-length DNA-A and DNA-B sequences of three of these divergent BGVs, and the DNA-A from a fourth virus, were determined from overlapped PCR products (see Methods). The isolated viruses exhibited the typical bipartite genome organization of New World BGVs, with a DNA-A containing four ORFs (*ACI- 4*) on the complementary-sense strand and one ORF (*AVI*) on the viral-sense strand; the DNA-B, on the other hand, displayed one ORF (*BVI*) in the virion-sense strand and another ORF (*BCI*) in the complementary strand. The complete DNA-A sequences of those viral isolates were analyzed by BLAST-N for preliminary species assignment based on the 89% threshold level established by the ICTV (Fauquet et al., 2008). As a result, it was determined that the four isolates were novel BGV species, and provisional names were assigned to them. The DNA-A of *Jacquemontia leaf distortion Yucatan virus* (JacLDYuV), isolated from a *Jacquemontia pentantha* (Convolvulaceae) plant (sample 534; Fig.1), was 2602 nt long and showed the highest overall sequence identity (69.5%) to *Sida yellow leaf curl virus* from Brazil (Accession No. EU710750). The

DNA-B was 2591 nt in length and shared a 190 nt segment (SI= 96%) with its cognate genomic component. A distinct BGV was isolated from a *Capraria biflora* (Scrophulariaceae) plant that exhibited shortening of internodes and yellow spotted leaves (sample YP#446; Fig.1). The DNA-A and DNA-B components of *Capraria yellow spot Yucatan virus* (CapYSYuV), were 2624-nt and 2584-nt in length, respectively, and shared a 154 nt common region displaying an unusually low sequence identity (i.e., 72%). The closest relative of CapYSYuV was *Rhynchosia mosaic Sinaloa virus* (DQ406672) with an overall DNA-A sequence identity at 81%. The DNA-A of *Abutilon golden mosaic Yucatan virus* (AbGMYuV) was isolated from an *Abutilon permolle* (Malvaceae) plant (sample YP#122; Fig.1). The DNA-A of this virus was 2629 nt in length, and exhibited the highest sequence identity (81%) with *Desmodium leaf distortion virus* (DesLDV; No. DQ875870). The fourth novel BGV was isolated from a *Vigna elegans* (Fabaceae) plant exhibiting yellow mosaics and slightly distorted leaves (sample 991; Fig.1), which was collected in Morelos state (see geographical coordinates in Methods). The DNA-A of *Vigna yellow mosaic virus* (ViYMV) was 2602 nt in length, and displayed the highest sequence identity (85%) with *Bean calico mosaic virus* (BCaMV; AF110189); the DNA-B molecule was 2564 nt in length, and shared a 192 nt sequence (SI=97%) with its cognate DNA-A.

Phylogenetic relationships of the isolated begomoviruses

A phylogenetic tree of the complete DNA-A sequences of ~50 New World BGVs as well as some selected bipartite Old World BGVs, was generated using the neighbor-joining method with 1,000 bootstraps replications; the tree was rooted with the sequence of *Maize streak virus* (MSV), the type species of the genus *Mastrevirus*. The resulting dendrogram (Figure 2) confirmed that the novel BGVs belong to the SLCV lineage. VYMV and

CapYSYuV grouped into a cluster of Mesoamerican BGVs that includes two well-studied BGVs: Cabbage leaf curl virus (CaLCuV) and Pepper golden mosaic virus (PepGMV); AbGMYuV grouped into a subclade that include both Mesoamerican and South American viruses, while JacLDYuV was basal to the clade. The phylogeny reconstructed from the full-length DNA-B sequences was not utterly congruent with the DNA-A dendrogram. For example, the DNA-B of JacLDYuV grouped within the AbMV clade, whereas CapYSYuV DNA-B was closest to the homologous component of the PHYVV cluster (Suppl. Fig.1). Incongruent phylogenies of the DNA-A and DNA-B components of a BGV are generally explained by recombination and/or reassortment events in the evolutionary history of that viral species (Briddon et al., 2010).

Recombination Analysis

Analysis of DNA-A components using the suite of programs for detection of recombinant breakpoints integrated in the RDP4 package (Martin et al., 2010) did not reveal significant evidence of recombination in CapYSYuV, JacLDYuV, and ViYMV (data not shown). In contrast, one recombination event with breakpoints at positions 1969 and 2595 was identified in AbGMYuV by several recombination detection methods. The recombinant 626 nt long segment was presumably derived from DesLDV (the “minor progenitor”), and encompasses the complete *AC4* ORF, the 5' half of the *AC1 (Rep)* gene and a part of the common region (Suppl. Fig 2). The remaining part of AbGMYuV DNA-A was derived from a close relative of *Corchorus yellow spot virus* (CoYSV). The recombination event between CoYSV and DesLDV was well supported by six methods in the RDP4 package (GENECONV, $p=2.649 \times 10^{-34}$; RDP, $p=3.606 \times 10^{-31}$; BootScan, $P=6.429 \times 10^{-27}$; MaxChi,

$p=1.172 \times 10^{-27}$; Chimaera, $p=1.636 \times 10^{-25}$; 3Seq, $p=7.617 \times 10^{-50}$). It is noteworthy that the hypothetical major progenitor of AbGMYuV does not belong to the SLCV clade.

Infectivity of cloned DNA-A and DNA-B components

Since all novel BGVs were isolated from doubly-infected plants, full-length clones of both DNA-A and DNA-B components of these viruses (with the exception of AbGMYuV) were generated to verify their capability to induce symptoms and multiply its genome in plants. Infectious clones of JacLDYuV, CapYSYuV and ViYMV were produced via PCR amplification by using back-to-back primers designed on unique restriction sites on each genomic component (see Methods). Biolistic inoculation of *Nicotiana benthamiana* plants with DNA-A and DNA-B clones of JacLDYuV produced evident symptoms at 10 dpi in all tested plants (8/8). Disease signs included strong shortening of internodes, yellowish and severe deformation of new leaves (Fig.3a). The symptoms did not decrease with time, and plants did not developed flowers even after 40 dpi. In contrast, 6 out 8 plants bombarded with the CapYSYuV clones developed mild symptoms including a slight delay in growth and mild leaf curling (Fig.3b), which lessened gradually after 18-20 dpi, indicating a process of infection remission. Finally, the clones of ViYMV produced symptoms in 6 out 8 inoculated plants, which exhibited mild chlorosis and deformation of new leaves 12 dpi, although remission of symptoms was also observed after 25 dpi (Fig. 3c). These observations demonstrated that the three inoculated BGVs have different capabilities to induce symptoms in the common experimental host (*N. benthamiana*), and confirmed the complementary nature of the isolated viral genomic components, which was an important issue in the case of CapYSYuV, whose common region exhibited poor conservation in

nucleotide sequence (72% SI) between cognate genomic components, although the relevant *Ori* elements were well preserved.

Identification of potential replication specificity determinants of the SLCV clade BGVs

The four isolated BGVs exhibited the characteristic iteron arrangement of the SL-C viruses although their iterative elements were not similar. Thus, CapYSYuV harbored six iterons with a TGGAGTCC consensus whereas AbGMYuV and JacLDYuV hold five iterons with a TGGTGTCC core sequence. In contrast, ViYMV exhibited five iterons with a TGGAGACC sequence, differing in the 6th nucleotide from the former elements. Sequence analysis of the remaining 26 SL-C members showed that, with the only exception of Rhynchosia mosaic Sinaloa virus (RhMSinV), all of them harbor iterons with either a TGGTGTCC (like JacLDYuV) or TGGAGTCC (like CapYSYuV) sequence. The single difference in the 4th nucleotide of the iterative sequence is critical for recognition by the cognate Rep protein, as established in experiments of pseudorecombination between two strains of Euphorbia mosaic virus (i.e., “Yucatan Peninsula” and “Jalisco”) that were unable to form viable reassortants in a common host, thus indicating replication incompatibility (Gregorio-Jorge et al, 2010). To identify potential DNA binding specificity determinants (SPDs) in the Rep proteins encoded by the SL-C members, an analysis of all of them by the CAGHIP method (Londoño et al., 2010) was carried out. As a result, four putative SPDs were identified in those proteins (Figure 3). These potential SPDs were situated at defined distances from the conserved RCR Motifs I and II, as formerly established in Rep proteins of three families of circular ssDNA viruses (Londoño et al., 2010). However, the positions of Rep SPDs relative to the initial residue in proteins

encoded by the SL-C members slightly differ from those observed in the homologous proteins of other NW and OW BGVs. Indeed, the Rep SPDs of most BGVs have been theoretically mapped at two discrete segments encompassing the residues 5, 8, 10, close to the RCR Motif I, and 69 and 71, in the vicinity of Motif II (Londoño et al., 2010, Mauricio-Castillo et al., 2014), whereas the putative Rep SPDs of SL-C viruses are placed at positions 9 and 11 (near to Motif I) and 71 and 73 (close to Motif II). As previously pointed out, most of the SL-C members exhibit iterons with either a TGGAGTCC or TGGTGTCC sequence, differing in a single nucleotide; accordingly, their Rep proteins display a single difference in the SPD at position 11, which invariably is Q in the former group, and is either A, T or S in the latter one. (Fig. 3). Interestingly, the only two viral species of the SLCV lineage exhibiting iterons different to the aforesaid ones, i.e., VYMV (TGGAGACC) and RhMSinV (TGGAGGAY), were also the only viruses encoding Reps that do not display an Asn residue at position 71. In effect, VYMV Rep has a Ser residue, and RhMSinV Rep an acidic residue (Asp) at that position. Those data are in strong accord with the hypothetical correlation between Rep SPDs and cognate iterons in geminiviruses, and suggest that the SPDs in the vicinity of Motif II might to influence the recognition of specific nucleotides at the iteron 3' end (Londoño et al., 2010).

Identification of a distinctive Rep domain of SLCV clade members.

The SL-C viruses were firstly distinguished from other BGVs because of their unusual arrangement of iterons (Arguello-Astorga et al., 1994a, b). To determine whether the 30 viral species grouped into the SLCV branch of the global dendrogram in Fig. 2, display additional distinctive features that could be used as hallmarks of this BGV lineage, all proteins encoded in the DNA-A component of those viruses (CP, Rep, TrAP, REn and

AC4) were systematically compared with those encoded by other New World BGVs. For this, the predicted proteins of the SL-C members were aligned with their homologs in other American BGVs, and the existence of distinctive segments or motifs between the compared BGV lineages was assessed. The analysis revealed SL-C specific motifs solely in the Rep and AC4 proteins. Indeed, several lineage-specific motifs were identified along the N-terminal segment (aa residues 1-160) of Rep but not in other of its domains (Fig. 4 and Supp. Fig 3). The extension of the comparative analysis to include Rep sequences of BGVs natives to the Old World, such as the Corchoviruses (Ha et al., 2006, 2008), Legumoviruses (Quazi et al., 2007), Sweepoviruses (Briddon et al., 2010), and other lineages of bipartite and monopartite BGVs from Africa, Madagascar, Europe, Asia, and Oceania, showed that the SLCV lineage-specific motifs are not present in proteins encoded by other members of the genus *Begomovirus* (Fig. 4, and data not shown). In view of those puzzling data, suggestive of an antique event of lateral DNA transfer between unrelated viral lineages, we expanded the comparative analysis to include the Rep proteins of all members of the family *Geminiviridae*, looking for a plausible ancestor of the Rep N-terminal domain of SL-C viruses. A phylogenetic tree based on an alignment of the N-terminal domain (1-160) of all geminivirus Rep sequences available at public databases was generated by Clustal W and the resulting dendrogram was subsequently used as a basis for more detailed comparisons. In Figure 5 a cladogram derived from the comparisons of partial Rep proteins of SLCV lineage BGVs with a number of geminiviruses representative from different genera and geographical lineages, is illustrated. Several data of this phylogeny are noteworthy: 1) the truncated Rep₁₆₀ proteins of the SL-C BGVs form a cluster clearly separated from the main branch encompassing the Rep N-terminal domain of NW and OW BGVs, topocuviruses (TPCTV), turncurtoviruses (TCTV), and the strains of BCTV, a curtovirus; 2) the Rep₁₆₀ of

two curtoviruses, Horseradish curly top virus (HrCTV) and Spinach severe curly top virus (SpiSCTV) were clearly related to those of the SL-C BGVs; 3) Rep₁₆₀ of JaCLDYuV (one of the novel BGVs characterized in this study) was basal to the SLCV branch, suggesting a closer relationship with the homologous protein domain of the SL-C common ancestor; 4) the only known member of the genus *Topocuvirus* have a Rep₁₆₀ practically indistinguishable from those of some NW BGVs; 5) The Rep₁₆₀ of TCTV, the only species in the genus *Turncurtovirus*, is closely related to Rep₁₆₀ of two BCTV strains, BSCTV and BMCTV. These incongruent relationships among the N-terminal of Rep proteins strongly suggest that intermolecular recombination between members of different lineages has taken place several times along the evolutionary history of the *Geminiviridae*. A significant but provisional conclusion derived from the last comparative analysis is that the N-terminal domain the Rep protein encoded by the SL-C viruses was not derived from a begomoviral ancestor but from a curtovirus related to HrCTV and SpiSCTV. However, this hypothesis has several shortcomings that will be discussed later.

Modeling of VYMV and BCaMV Rep proteins

Since VYMV harbors iterons with a nucleotide sequence that is unique among the SL-C BGVs, a comparison of the predicted 3D structure of VYMV Rep and the homologous protein of its closest relative, *Bean calico mosaic virus* (BCaMV), that exhibits iterons differing in two nucleotides (TGGTGTCC) might provide important clues about the Rep aa residues determining its DNA-binding specificity, as recently showed in the case of two related BGVs from Sinaloa, Mexico, that exhibit divergent iterons ([Mauricio-Castillo et al., 2014](#)). Accordingly, 3D models of the DNA-binding domain of VYMV and BCaMV were built based on the solved NMR structure of TYLCSV Rep 4-121 (see Methods). The

structural model of VYMV Rep1-120 (Figure 6) comprise a core of five β -strands arranged in a 5-stranded antiparallel β -sheet, surrounded by two α -helices, and two small 2-stranded β -sheets. The β -strands 2 and 4 of the central sheet include the so-called RCR motifs I and II, and their extensions, the β -1 and β -5 strands, form a small antiparallel β -sheet that has been involved in specific DNA binding (Campos-Olivas et al. 2001). Comparisons of the modeled Rep1-120 structures of VYMV and BCaMV revealed clear differences in the region encompassing the β -1/ β -5 sheet. As can be appreciated in the modeled protein, VYMV Rep exhibits two polar amino acids, one with a large side chain (Q11) and other with a small one (S71), protruding from the former structural region, while BCaMV Rep has in the equivalent positions to T11 and N71, respectively. It is important to point out that VYMV and Rhynchosia mosaic Sinaloa virus (RhMSinV) are the only SL-C members whose Rep protein displays an amino acid residue at position 71 which is not asparagine (N). Indeed, VYMV Rep display serine (S) at that position, whereas RhMSinV has an acidic residue, aspartate (D); these differences are in close accord with the observation that the latter two BGVs harbor iterons with a unique sequence. In conclusion, the amino acid in the β -1/ β -5 sheet of VYMV and BCaMV Rep proteins are notoriously different, and expose distinct side chains toward the hydrophilic surface of the mini β -sheet, thus providing dissimilar electrostatic and hydrogen-bonding patterns for specific interactions with the DNA, which could account for the affinity of those proteins by different iterons.

Analysis of the distinctive AC4 protein of SLCV lineage viruses.

Since that the AC4 ORF is contained entirely within the region encoding the Rep domain 1-160, it may be anticipated that AC4 proteins of SL-C members would be significantly

divergent in sequence from their homologs encoded by other geminiviruses. Indeed, the alignment of the former proteins with the AC4/C4 proteins of diverse BGVs revealed a very low aa sequence identity between them, ranging of 28 to 15% (Supp. Figure 4). Furthermore, the predicted AC4 proteins of SL-C BGVs were consistently larger than their homologous in other geminiviruses (with the exception of the curtoviruses HrCTV and SpiSCTV), with ~120 aa residues in length, that is in contrast with the small size (~85 aa) of the predicted AC4/C4 proteins of most NW and OW BGVs, topocoviruses, turncurtoviruses and the curtovirus BCTV(data not shown). The uncommon size and amino acid sequence of the predicted AC4 proteins of SL-C viruses is another distinctive molecular characteristic of this atypical lineage of NW BGVs, and suggest functional differences with their positional analogous in other geminiviral lineages.

DISCUSSION

It is generally assumed that the begomoviruses evolved in the Cretaceous period, prior to the tectonic breakup of supercontinent Gondwana into the landmasses of Africa, India, South America, Antarctica, and Australia, which took place 120-110 MYA (...). This hypothesis is based on the observation that phylogenies generated from alignments of the complete DNA-A sequences of BGVs consistently generate continental lineages, suggesting that those viruses evolved in geographic isolation for extensive periods. Thus, the BGVs from the Old World segregate into clusters originating in the Indian subcontinent, Asia, Japan and Africa (Briddon et al, 2010). The New World BGVs form a separate major branch, with several distinct clades that are not restricted to specific regions of the Americas. The prevalent view is that the BGVs were introduced to the American continent by the region of Alaska (Lefevre et al., 2011; ...), although an alternative

scenario has been recently proposed, that hypothesize that the BGVs arrived to South America when a land connection with the Antarctic continent existed, and the climatic conditions in the latter supported a subtropical vegetation and populations of whiteflies (Cardenas-Conejo, 2011; Cardenas-Conejo and Arguello-Astorga, in preparation). In any case, an evolutionary radiation of BGVS during its spread across the american continent took place, giving rise to secondary lineages, such as the clades of the AbMV, -BGMV,- SLCV and others defined in different phylogenetic studies. In this work we have established, through molecular characterization of new species of the SLCV clade and comprehensive comparative sequence analyses, that this lineage is different from other New World BGVs in several molecular traits associated with viral replication and relevant domains of the Rep and AC4 proteins.

METHODS

Recombination analysis

The recombination analysis was performed using the recombination detection methods included in the RDP 4.16 package with default settings (Martin, et al., 2010). The analysis was based on alignment of the full-length DNA-A components of AbGMYuV, CoYSV and DesLDV that was performed with Clustal V algorithm of the MegAlign application of Lasergene (DNASTar Inc., Madison, WI, USA).

Homology structural modeling

Homology models for monomeric Rep proteins of VYMV and BCaMV were built using I-TASSER server, which employs a fragment-based method in which fragments are excised

from multiple template structures and reassembled, based on threading alignments (Roy et al., 2010) The NMR solution structure of Rep protein from a begomovirus (TYLSCV) where specified as a template with target-template alignment to the I-TASSER server (PDB accession code 1L2M). One model for each virus was obtained with a C-score of 1.98 for both the VYMV and BCaMV. The structure of each model was further optimized to eliminate bad torsion angles, Van der Waals clashes or total energy using the repair option of FoldX (Guerois et al., 2002). The quality of the models was analyzed using the Molprobit server and ProSa server (Chen et al., 2010; Wiederstein et al., 2007). The stereochemistry analysis shows that 88% and 89% of the peptide bonds were in the favored regions of the Ramachandran plot, respectively for the VYMV and BCaMV models. The ProSa z-scores values for the VYMV and BCaMV models were -4.7 and -4.5 respectively; those values are in the range of scores typically found in proteins of similar sequence length. UCSF Chimera was used for molecular graphics analyses and imaging (Pettersen et al., 2004).

FIGURE LEGENDS

Figure 2. DNA-A inferred evolutionary relationships of viruses isolated in this work compared with selected begomoviruses. Phylogenetic tree indicating the relationship between the DNA-A genomic component of viruses isolated in this work (red caption) and selected begomoviruses is shown. The SLCV clade members are denoted with a red triangle depicted before the name. MSV mastrevirus sequence was used as outgroup. Names and GenBank access numbers for each virus sequence used are listed in supplementary Table 1. DNA-A phylogeny was inferred with a total of 2235 positions in the final dataset using the Neighbor-Joining method (Saitou and Nei, 1987). Bootstrap

values equal or above a 60% significance threshold, based on 1000 replicates, are depicted in each node as follows: from 60-79% (○), from 80-89% (●) and from 90-100% (■). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the number of differences method (Nei and Kumar, 2000) and are in the units of the number of base differences per sequence. The analysis involved 52 nucleotide sequences. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA5 (Tamura et al., 2011).

Supplementary online resources

Supp. Figure X. DNA-B inferred evolutionary relationships of viruses isolated in this work compared with selected begomoviruses. Phylogenetic tree indicating the relationship between the DNA-B genomic component of viruses isolated in this work (red caption) and selected begomoviruses is shown. The SLCV clade members are denoted with a red triangle depicted before the name. MSV mastrevirus sequence was used as outgroup. Names and GenBank access numbers for each virus sequence used are listed in Supp. Table 1. DNA-B phylogeny was inferred with a total of 1470 positions in the final dataset using the Neighbor-Joining method. Bootstrap values equal or above a 60% significance threshold, based on 1000 replicates, are depicted in each node as follows: from 60-79% (○), from 80-89% (●) and from 90-100% (■). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the number of

differences method (Nei and Kumar, 2000) and are in the units of the number of base differences per sequence.

Suppl. Figure X2. Recombination event in the AbGMYuV DNA-A detected by RDP 4. The schematic representation of AbGMYuV genomic component A is at the top of the figure and the RDP pair-wise similarity plot is shown below. The blue region indicates the major parent section (CoYSV) and the red region the minor parent (DesLDV). The color lines in the similarity plot indicate the comparison between two virus sequences. The beginning breakpoint (1969 nt) and the ending breakpoint (2595 nt) are indicated by the vertical lines.

Suppl. Fig. X3 AC4 protein phylogeny. Phylogenetic tree indicating the relationship between the AC4 protein of viruses isolated in this work (red caption) and AC4 proteins of selected begomoviruses is shown. The SLCV clade members are denoted with a red triangle depicted before the name. Curtoviruses (BMCTV, BSCTV, HrCTV, SpiSCTV), and the single members of the genera *Turncurtovirus* (TCTV) and *Topocovirus* (TPCTV) AC4 proteins are denoted with a green circle and a purple square depicted before the name, respectively. GrapevineGV amino acid sequence was used as outgroup. AC4 phylogeny was inferred with a total of 57 positions in the final dataset using the Neighbor-Joining method. Bootstrap values equal or above a 60% significance threshold, based on 1000 replicates, are depicted in each node as follows: from 60-79% (○), from 80-89% (●) and from 90-100% (■). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the number of differences method, and are in the units of the number of base differences per sequence.

FIGURES

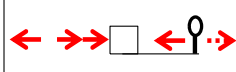

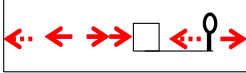

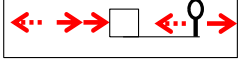

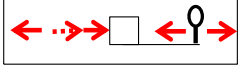

Provisional name	Host	Genomic components	Common Region	Closest relative	Iteron arrangement	
Jacquemontia leaf deformation Yucatan virus (JacLDYuV)	<i>Jacquemontia pentantha</i> (Convolvulaceae)	DNA-A (2602 nt) DNA-B (2591 nt)	190 nt (SI= 96%)	SiYLCV (EU710750) SI= 69.5%		
Capraria yellow spot Yucatan virus (CapYSYuV)	<i>Capraria biflora</i> (Scrophulariaceae)	DNA-A (2624 nt) DNA-B (2584 nt)	154 nt (SI= 72%)	RhMSinV (DQ406672) SI= 81%		
Abutilon golden mosaic Yucatan virus (AbGMYuV)	<i>Abutilon permolle</i> (Malvaceae)	DNA-A (2629 nt)	ND	DeLDV (DQ875870) SI= 81%		
Vigna yellow mosaic virus (ViYMV)	<i>Vigna elegans</i> (Fabaceae)	DNA-A (2602 nt) DNA-B (2564 nt)	192 nt (SI= 97%)	BCaMV (AF110189) SI= 82%		

Figure 1

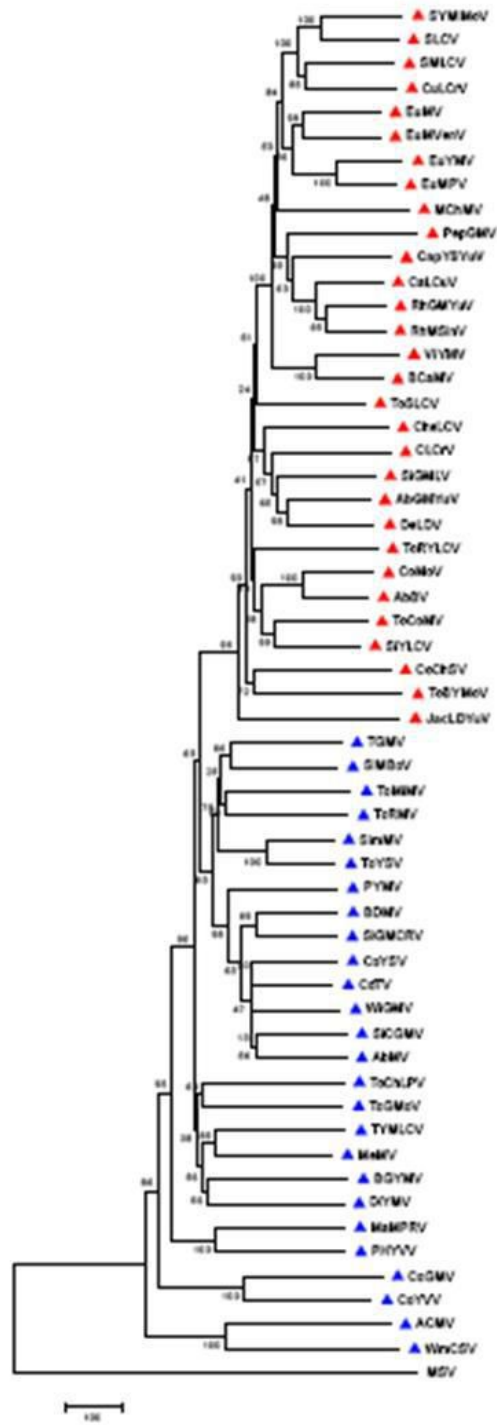


Figure 2

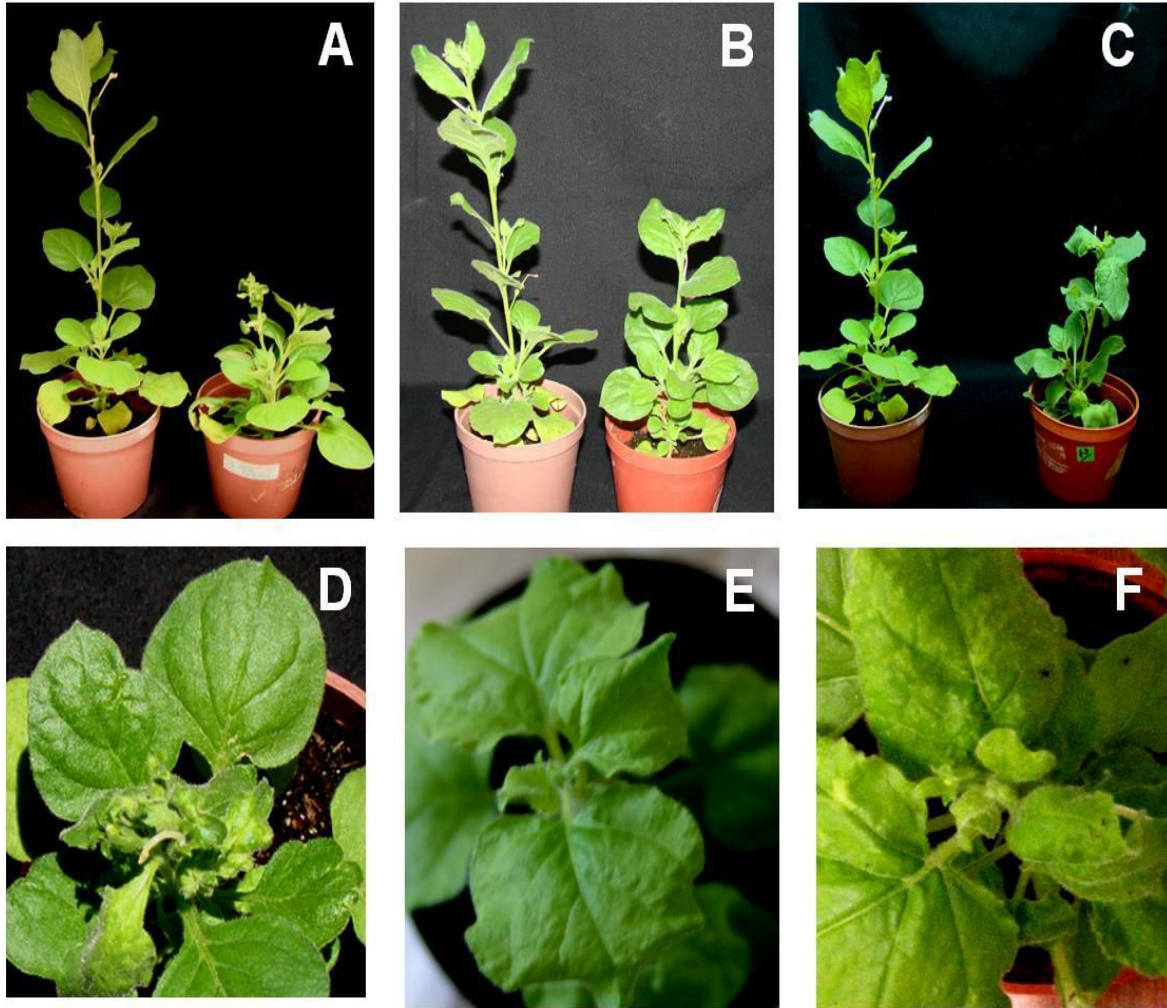


Figure 3

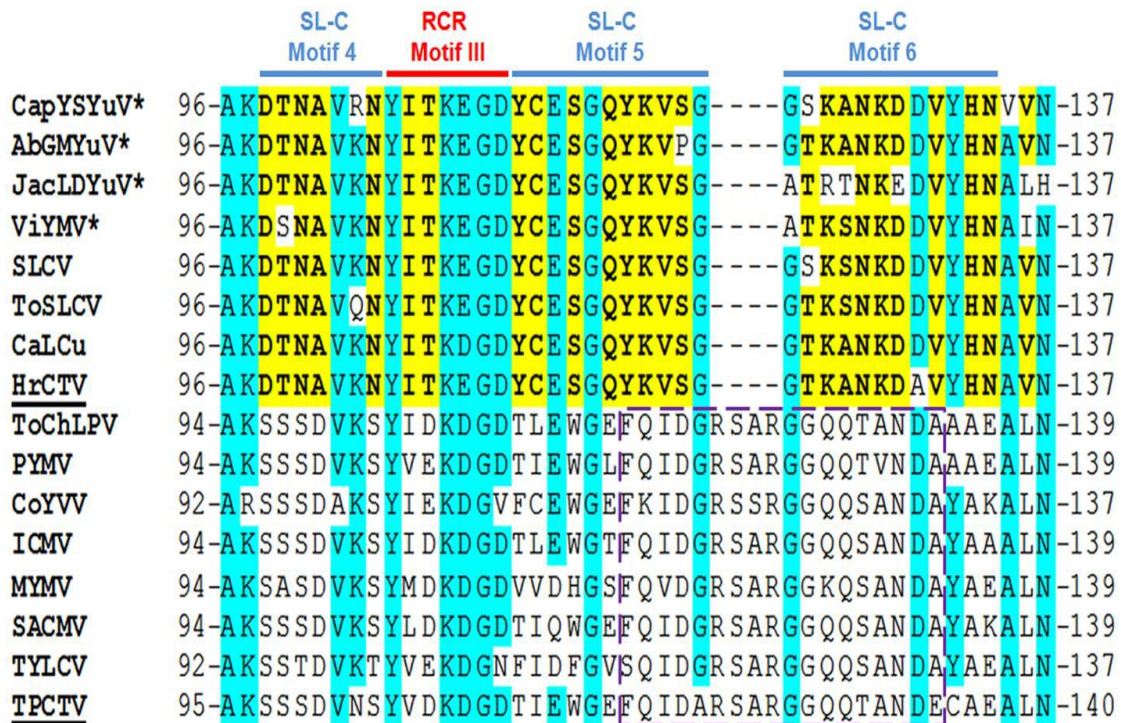


Figure 4

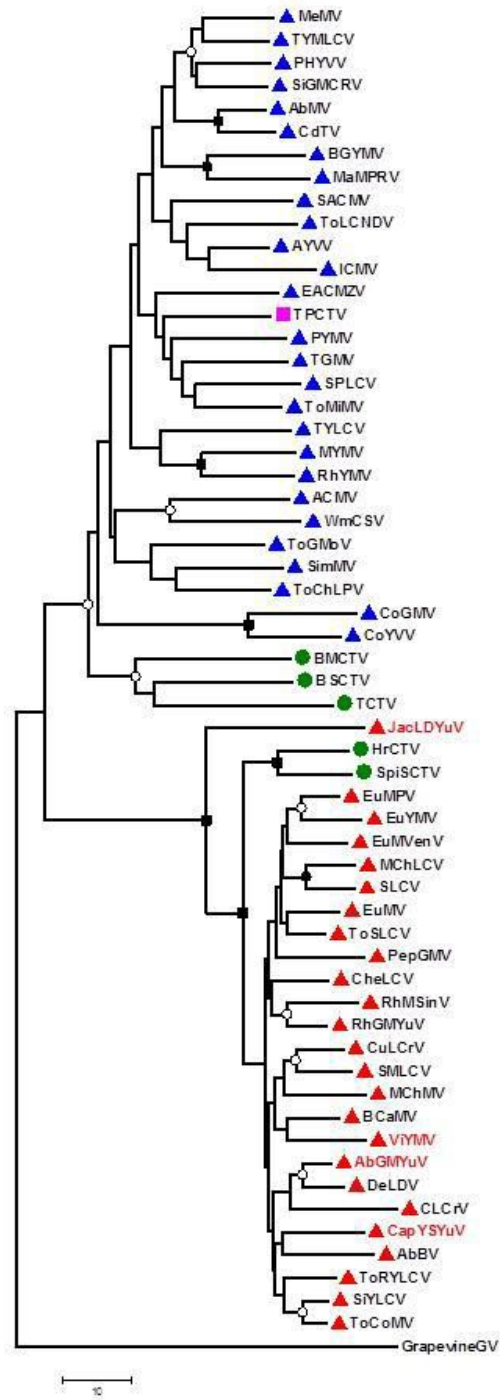


Figure 5

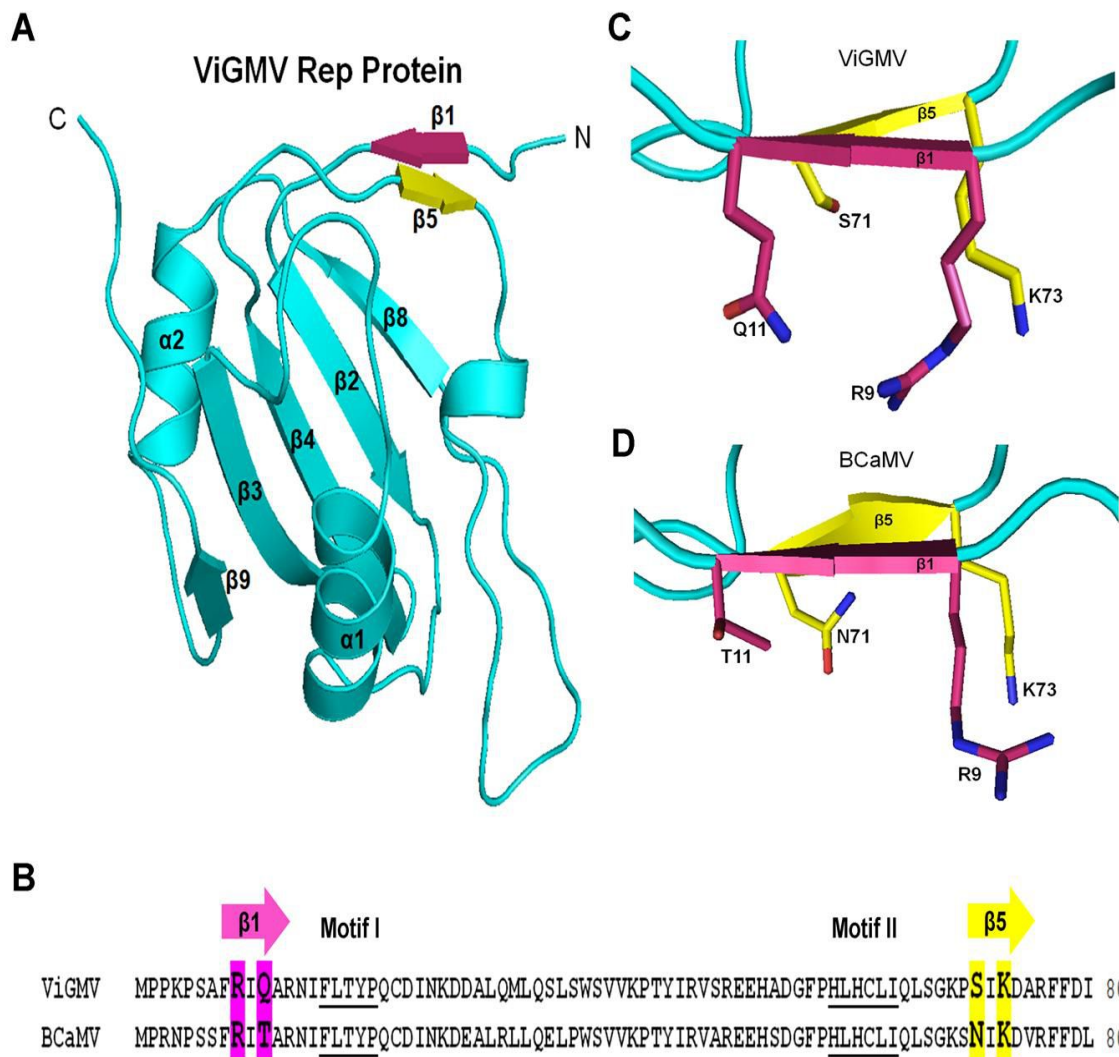


Figure 6

	SPDs-R1		Motif I		Motif II	SPDs-R2					
ITERON	MPRAPGS	FRLA	AKNI	FLTYPQC	---36---	HLHCLL	QLSGKS	NIK	DARFFD	AbGMYuV	
	MPRQPNK	FRLA	SRNI	FLTYPQC	---36---	HLHILI	QLSGKS	NIK	DERFFD	JacLDYuV	
	MPRNPSS	FRLT	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DVRFFD	BCaMV	
	MPRNPKS	FRLA	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	CaLCuV	
	MPRNPNS	FRLS	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	CuLCrV	
	MPLKPKT	FRLA	ARNI	FLTYPQC	---36---	HLHCLL	QLSGKS	NIK	DARFFD	DeLDV	
	MPRNPNS	FRLT	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	EuMPV	
	MPRNPNS	FRLA	AKSI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	EuMV-YP	
	MPRNPNS	FRLT	AKNI	FLTYPRC	---36---	HLHCLI	QLSGKS	NIK	DARFFN	EuYMV	
	TGGTGTCC	MPQNPKS	FRLS	ARNI	FLTYPKC	---36---	HLHCLV	QLSGKS	NIK	DAGFFD	MeChLCuV
		MPRSPKS	FRVA	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	NARFFD	MeChMV
		MPREPKS	FRLA	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	NARFFD	SiGMLV
		MPRNPQK	FRLA	AKNI	FLTYPQC	---36---	HLHCLV	QLSGKS	NIK	DVRFFD	SiYLCV
		MPRNPNS	FRLT	ARNI	FLTYPRC	---36---	HLHCLI	QLSGKS	NIK	DAGFFD	SLCV
		MPRNPNS	FRLT	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKT	NIK	NARFFD	SMLCV
		MPRNPNI	FRLT	ARNI	FLTYPQC	---36---	HLHCLV	QLSGKS	NIK	DVRFFD	ToCoMV
		MPRNPNQ	FRLA	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	ToRYLCV
		MPRNPNL	FRLA	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	ToSLCV
		MPRQNKs	FRLS	SKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DCRFFD	SpSCTV
	TGGAGTCC	MPRNPNS	FRLQ	ANNI	FLTYAQC	---36---	HLHCLL	QLSGKS	NIK	DAGFFD	CaYSYuV
MPRLPHS		FRLQ	ARNI	FLTYPRC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	AbMV	
MPRNPNS		FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	CheLCV	
MPRNPDS		FRLQ	ARHI	FLTYPKC	---36---	HLHCLI	QLSGKS	NIK	NKRFFD	CLCrV	
MPRLPNS		FRLQ	ARNI	FLTYPRC	---36---	HLHCLV	QLSGKS	NIK	DARFFD	CoMoV	
MPRQPKL		LRVQ	SKNI	FVTYPQC	---36---	HLHCLV	QLSGKP	NIK	DVRFFD	CoChSV	
MPRNPNS		FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	EuMV-Jal	
MPRNPNS		FRLQ	AKNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	EuMVenV	
MPLPPKS		FRLQ	CKNI	FLTYPQC	---36---	HLHCLI	QLTGKC	NIK	DARFFD	PepGMV	
MPRNPNS		FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	RhGMYuV	
MPRQPNT		FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DARFFD	TBYMV*	
MPRQPNs		FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKS	NIK	DCRFFD	HrCTV	
TGGAGACC		MPPKPSA	FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGKP	SIK	DARFFD	ViYMV
TGGAGGAC	MPRNPNS	FRLQ	ARNI	FLTYPQC	---36---	HLHCLI	QLSGES	DIK	NARFFD	RhMSinV	

Figure 7

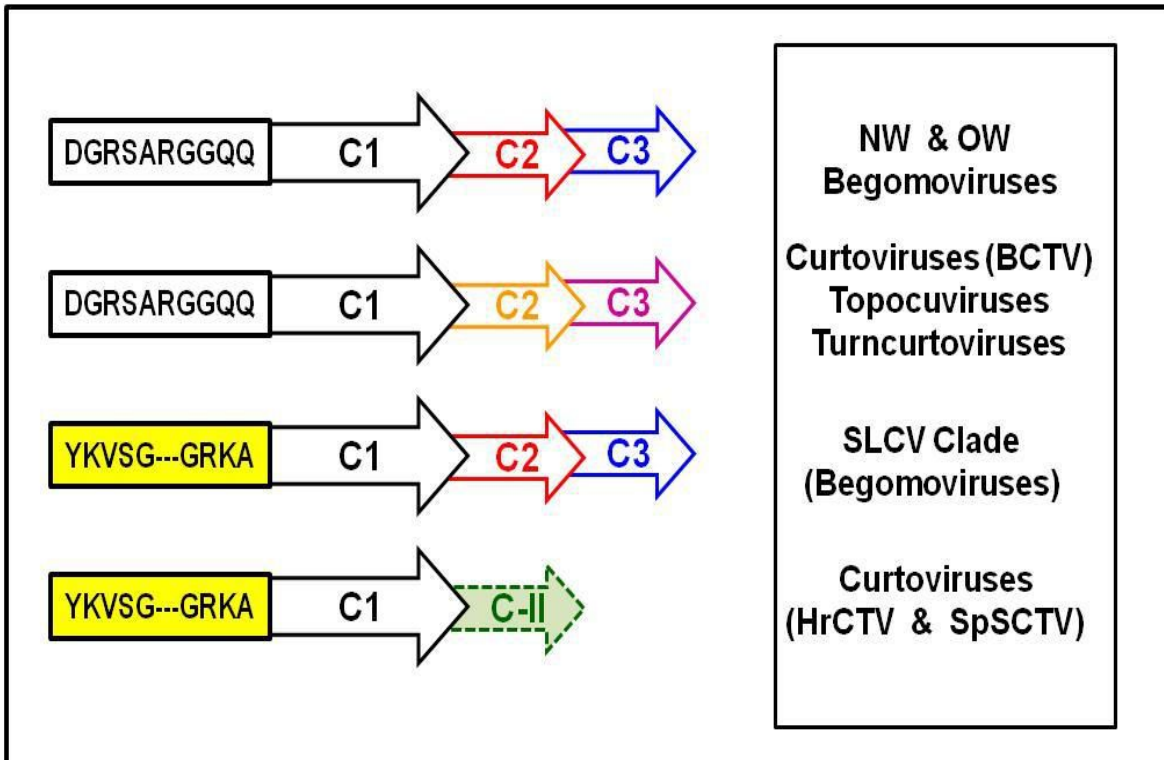
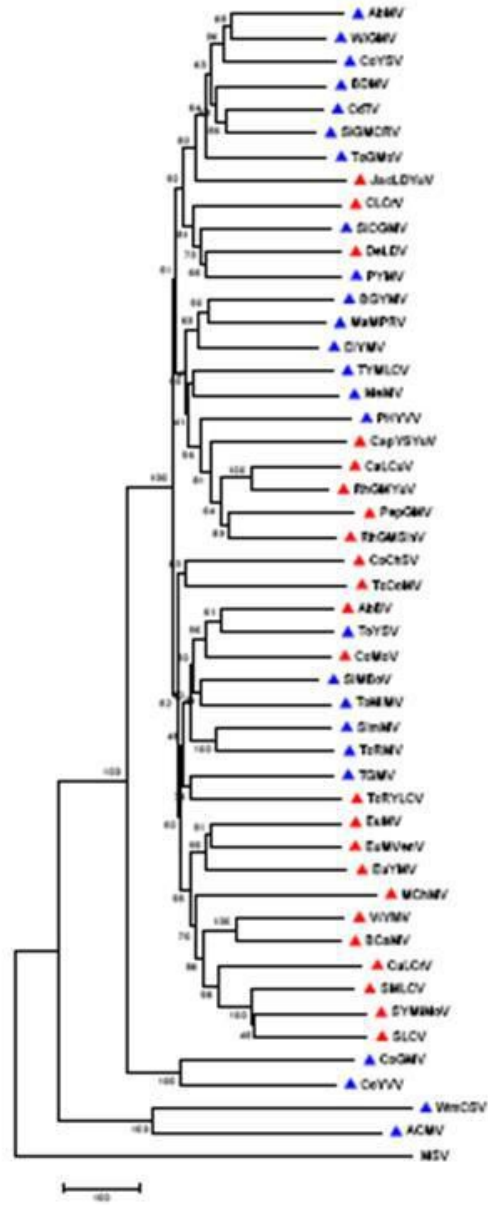
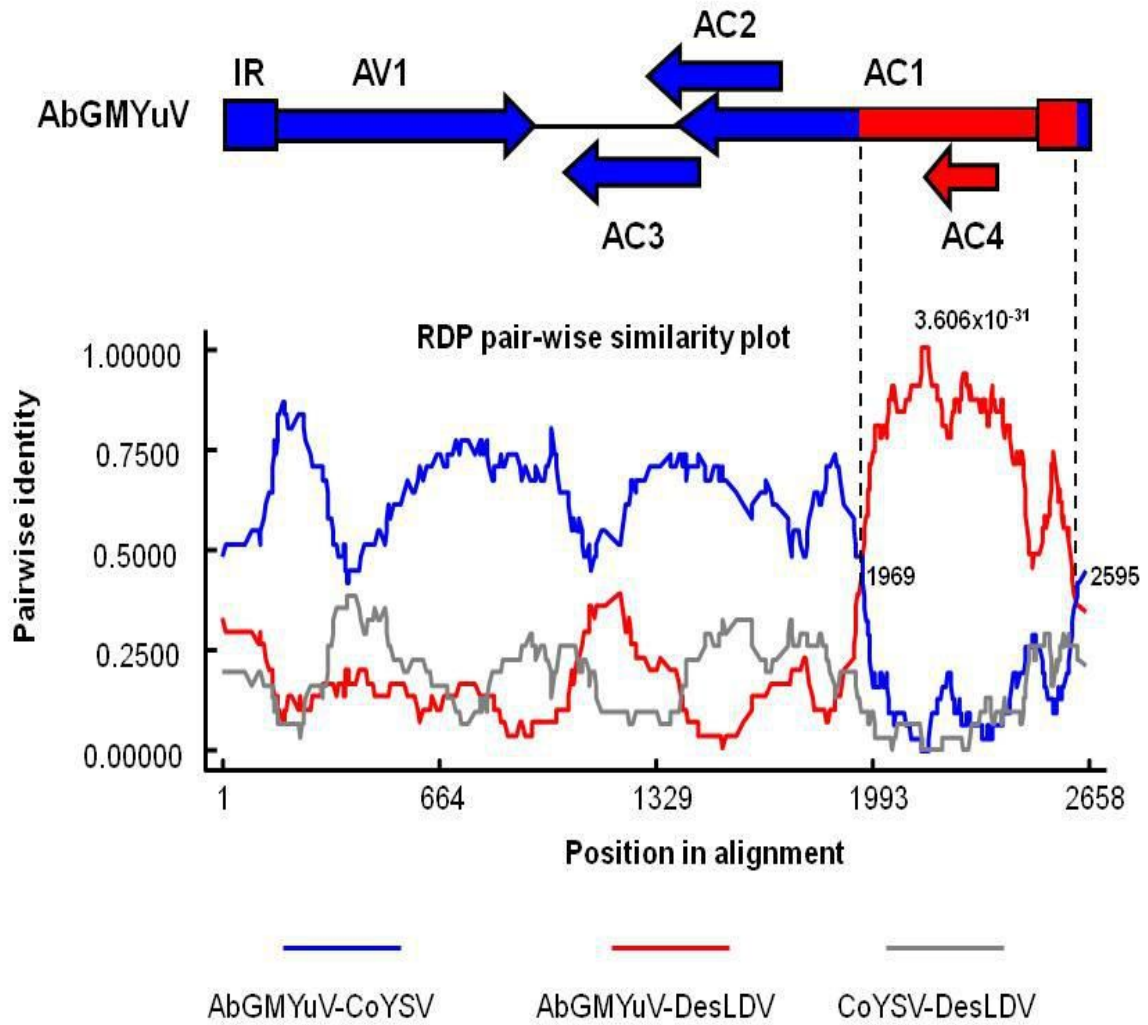


Figure 8

Supplementary Figures



Supp. Figure 1



Supp. Figure 2

Anexo 3

(Manuscript in preparation)

***Squash leaf curl virus* does not induce symptoms in tomato but increases symptoms severity in mixed infections with TYLCV.**

Iliana Torres-Herrera¹, Josefina León-Felix², Clara Monreal-Vargas³, Salvador Ambriz-Granados¹, Gerardo R. Argüello-Astorga¹.

¹Instituto Potosino de Investigación Científica y Tecnológica, A.C., Camino a la Presa de San José, C.P. 78216. San Luís Potosí, SLP, México. ² Centro de Investigación en Alimentación y Desarrollo, A.C. Apdo. Postal 32-A, Culiacán, Sinaloa, C. P. 80129.

& Author for correspondence. Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), Camino a la Presa de San Jose 2055, San Luis Potosí, México. C.P 78216. Phone: +52 (444) 8342000 Ext. 2079. Fax: +52 (444) 8342010. E-mail: grarguel@ipicyt.edu.mx

Key words: geminiviruses, recombination, mixed infection

Abstract

A survey of tomato commercial fields in Sinaloa, Mexico, revealed a high incidence of *Tomato yellow leaf curl virus-Israel* (TYLCV-IL), an Old World monopartite begomovirus that was recently introduced to the American continent. Analysis of 53 field samples revealed the presence of TYLCV in 41 symptomatic tomato plants. In addition, DNA sequences from *Squash leaf curl virus* (SLCV), a cucurbit-infecting begomovirus, were amplified from six TYLCV-infected tomato plants. Infectious clones of SLCV and TYLCV were obtained from one of those plants and tested in several host plants. Biolistic inoculation of SLCV DNA-A and DNA-B produced severe symptoms in both squash and *Nicotiana benthamiana* plants, but did not induce symptoms in tomato plants even though viral DNA could be amplified from young tissues, hence indicating poor adaptation of SLCV to this host. In contrast, co-inoculation experiments revealed that SLCV significantly contributed to symptoms development in mixed infections, because they were consistently more severe in tomato plants infected with both begomoviruses than those only infected by TYLCV. These observations indicate that exotic begomoviruses like TYLCV might assist the colonizing of new hosts by endemic begomoviruses, thus creating conditions for evolution of viral variants better adapted to plant species that are not their natural hosts.

Introduction

In Mexico, the state of Sinaloa is the largest producer of tomato (*Solanum lycopersicum*) for domestic consumption and export. In the valleys of Sinaloa, where this crop is grown, there have been huge economic losses due to diseases caused by RNA and DNA viruses. In recent years, whitefly-transmitted geminiviruses of the genus *Begomovirus* have been the main agents of those losses (Gómez-Jiménez et al. 2009). Several begomoviruses native of the New World have been identified in Sinaloa tomato crops along the last two decades, including to Tomato leaf curl Sinaloa virus, Chino del tomate virus (CdTV), Pepper huasteco yellow vein virus (PHYVV), Pepper golden mosaic virus (PepGMV) and mixed infections among them. More recently an Old World begomovirus, *Tomato yellow leaf curl virus* (TYLCV) has also being found affecting tomato crops in the Culiacan and Guasave counties of Sinaloa (Brown et al. 2002 Gómez-Jiménez et al. 2009). In this region, besides tomato, other commercially valuable species such as pepper, potatoes and cucurbits (squash, melon and watermelon) are also usually cultivated. In cucurbit have been detected begomoviruses as *Squash leaf curl virus* (SLCV), *Squash mild leaf curl virus*, *Cucurbit leaf crumple virus* (CuLCrV) and *Melon chlorotic leaf curl virus* (MCLCuV) (Idris et al. 2008). The phylogenetic analysis of New World begomoviruses recognizes an atypical lineage known as the *Squash leaf curl virus* (SLCV) clade (Arguello-Astorga et al., 1994; Padidam et al. 1995). SLCV differ from “typical” begomoviruses in the origin of replication with a different number and arrangement of iterons, and its Rep protein lacks several conserved motifs in the N-terminus of its homologs in other begomoviruses and has an AC4 gene that encodes a protein of 120 amino acids that is not

homologous to the “typical” C4 protein (Argüello-Astorga et al. 1994a, 1994b; Argüello-Astorga and Ruíz-Medrano 2001). Such agroecosystems promotes the increase of whitefly populations and their migration from one species to another. This behavior of the whitefly, coupled with the existing diversity of begomoviruses favors mixed infections, which cause the emergence of variants and even new begomoviruses due to recombination processes. In addition, the migration of begomovirus-infected whiteflies biotypes from other regions, and the evolution of the insect vector-begomovirus relationship may promote the introduction of begomoviruses in tomato plants (Rojas and Gilbertson, 2008). There are several strategies for determining diversity of geminiviruses in plant samples from agricultural regions as those reported by Mauricio-Castillo et al. (2006, 2007) and Cárdenas-Conejo et al. (2010). On the basis of the foregoing, the aim of this study was to detect mixed infections and characterize some of them in tomato crops in the valley of Culiacan, Sinaloa, Mexico.

Materials and Methods

Plant samples

Sample collection was carried out in two tomato fields located in the Valley of Culiacan, Sinaloa, Mexico (geographical coordinates: Field A, 24° 77' 12" 27"-107° 52' 31" 93"; Field B, 24 ° 46' 07" 09" -107 ° 34' 17" 05"). The expression of severe or mild symptoms of virosis, besides the presence of whitefly populations in the fields, were the criteria for sample gathering, although several asymptomatic plants were also collected. Fifty three samples from two commercial varieties of

tomato, “Saladette” and “Bola”, as well as some weeds found within the field were examined. The sampling was carried out in October 2006 and April 2007.

PCR-based characterization of begomoviruses

Total DNA was extracted of plant samples using a modified version of the Dellaporta method [Ref]. The presence of begomoviruses was assessed by using the total DNA extracts in PCR reactions with degenerated primers designed to amplify two overlapping genomic segments encompassing either the complete DNA-A or DNA-B molecules of begomoviruses. For the molecular characterization of the complete DNA-A of typical begomoviruses two pairs of primers, repDGR-fw/cpYMAC-rv, and repDGR-rv/CP70-fw were used (Mauricio et al., 2007). The complete DNA-A component of “atypical” begomoviruses belonging to the SLCV clade was assembled from overlapped viral amplicons obtained of PCR directed by two sets of primers specific for this lineage, SL2150-fw/ cpYMAC-rv and SL2150-rv/CP70-fw (Gregorio-Jorge et al. 2010). Two pairs of degenerated primers, BC1-290fw/ BV1-470rv, and BC1-290rv/ BV1-310fw were utilized to direct the amplification of begomoviral DNA-B sequences [Gregorio-Jorge et al. 2010]. PCR reactions were carried out in a final volume of 25 μ L, consisting in the following: Taq DNA polymerase reaction buffer 1X, $MgCl_2$ 1.5 mM, dNTPs 0.2 mM, oligonucleotides 1 μ M, Taq DNA polymerase 1 U (Taq Platinum, Invitrogen). Amplifications were done in a thermocycler (Applied Biosystems, model 2720) with the following conditions: initial denaturing 94°C/2 min, and then 35 cycles of 94°C/1 min, 55°C/1 min, 72°C/1 min; final extension 72°C/5 min. The viral amplicons were

cloned into the pGEM-T Easy vector (Promega) and subjected to Restriction Fragment Length Polymorphism (RFLP) analysis with EcoRI and HinfI enzymes. The products of digestion were separated by electrophoresis in 2.5% agarose gels and PCR clones with different restriction patterns were sequenced. Automated sequencing was carried out at LANBAMA, IPICYT (San Luis Potosí, Mexico) using a 3130 Genetic Analyzer (Applied Biosystems). Assembling of the complete genome sequence was carried out with the suite of applications of the Lasergene package (DNASTAR Inc., Madison, WI, USA).

Generation of infectious clones of TYLCV and SLCV

To clone the complete genome of TYLCV and the full-length genomic components A and B of SLCV, the total DNA extract of one tomato plant infected with both begomoviruses was used as template for the amplification by PCR of the viral genome segments. The TYLCV genome was obtained using abutted divergent primers designed over the unique BamHI site (i.e., TYL-BamHI-fw [ATAGGATCCACTTCTAAATGAATTTCC] and TYL-BamHI-rev [AATGGATCCCACATAGTGCAAGAC]). The generated ~2.7 Kb amplicon was cloned into the pBlueScript II vector. To obtain the full-length DNA-A and DNA-B molecules of SLCV abutted divergent primers were designed over unique restriction enzymes sites. The primers SLCVA SacI-fw (AATGAGCTCAAACGGAGGAAGCC) and SLCVA SacI-rev (TTTGAGCTCACTTCGTTTCACGTCAG) were used for DNA-A amplification, whereas the complete DNA-B was amplified using the primers SLCVB XbaI-fw (CAGTCTAGATGATGATGCGGATTTGACC) and SLCVB XbaI-rev

(CAGTCTAGATTCGGGATAAGAGGGCCCAATTG). In both cases, the produced ~2.6 Kb amplicons were cloned into pBlueScript II. Two independent clones of each SLCV genomic component were sequenced to confirm the absence of mutations, and one of them was selected to perform assays of infectivity in planta.

Plant infection assays

Plants of tomato (*S. lycopersicum* var. Saladette), *Nicotiana benthamiana* and squash (*Cucurbita pepo*) were inoculated using a low-pressure biolistic device. The leaves of young plants (third- to four-leaf stage) were directly shot at 120 psi helium pressure with tungsten particles (0.7 μm , BioRad, Hercules, CA) covered with the cloned viral DNA (1 μg each genome component). The biolistic experiments were carried out either as single or mixed inoculations. The infectious clone of TYLCV was digested with *Bam* HI, the SLCV-A clone with *Sac* I and that of SLCV B with *Xba* I, before inoculation to release the viral DNA from the cloning vector. Six tomato plants in a 4-6 leaf-stage were inoculated with the following DNA combinations: a) TYLCV, b) SLCV components A and B, c) TYLCV +SLCV components A and B. As control, tomato plants were inoculated with the empty cloning vector (pBluescript). After inoculation the plants were maintained in an insect-free growth chamber (27° C, 16 h light-8 h dark) and observed on a daily basis for the appearance of symptoms. All plants were examined by PCR at the end of the experiment (24-28 dpi). Three independent experiments of plant infectivity were carried out.

Results

Identification of begomoviruses on infected plants

During October 2006 and April 2007 a survey of two tomato commercial fields in the Valley of Culiacan was undertaken. These fields were infested with the whitefly *Bemisia tabaci* and a high proportion of the tomato plants exhibited symptoms of virosis such as leaf curling, yellowing, vein chlorosis and stunted growth. Leaf samples of a number of symptomatic and asymptomatic tomato plants were collected, and total DNA extracts from those samples were obtained and tested for the presence of begomoviruses by the PCR technique using several pairs of degenerated primers specific for this group of viruses (see Materials and Methods). Forty one from the 57 examined samples were PCR-positive for begomoviruses. The viral amplicons derived from these 41 plants with the pair of primers repDGR-fw and cpYMAC-rv were ~1530 nt in length, significantly larger than the 1400-1420 bp long amplicons expected for New World begomoviruses. Only in one sample an additional PCR product of ~1420 bp was observed, suggesting a mixed infection (Fig.1a). All viral amplicons were cloned (see Methods) and the recombinant plasmids were subjected to RFLP analysis using the *EcoRI* and *HinfI* endonucleases. All clones with the ~1530 bp amplicons produced an identical restriction fragment pattern, that was different to the one observed for clones with the ~1420 PCR product. Sequencing of the two different kind of amplicons showed that the largest one correspond to TYLCV-IL and the second one to PHYVV, two begomoviruses previously reported in tomato crops from Sinaloa.

The 41 samples containing TYLCV were additionally examined with a distinct pair of primers, namely, SL2150-fw and cpYMAC (see Methods), that direct the

amplification by PCR of a DNA-A segment of begomoviruses belonging to the *Squash leaf curl virus* (SLCV) clade, that are not detectable with the pair repDGR-fw/ cpYMAC-rv previously used. Eight samples produced a ~1350 bp long amplicon, thus indicating the presence of at least one additional begomovirus in each of those TYLCV-infected plants. RFLP analysis of cloned PCR products revealed two different restriction fragment patterns (Fig. 2b). Sequencing and ensuing BlastN analysis of the amplicon sequences showed that they corresponded to SLCV and *Pepper golden mosaic virus* (PepGMV), in six and two samples, respectively. The finding of SLCV in naturally infected tomato plants was unexpected because the known host range of SLCV only include plants belonging to the family Cucurbitaceae (e.g., melon, watermelon, squash and zucchini). To determine whether the SLCV isolate of tomato displays significant sequence differences with previously described SLCV isolates, we proceed to obtain the complete genome sequence of the former begomovirus. The DNA-A molecule of this virus consisted of 2,600 nt displaying the highest level of nucleotide sequence identity (SI) to SLCV-[USA:Cal:] (accession no. DQ285016.) at 9X.X%, while the DNA-B molecule was 2,xxx nt in length displaying the highest identity (SI= 9X%) to the homologous genome segment of SLCV.

Infectivity of clones of TYLCV, SLCV and mixed infections in tomato plants

Since the previously known host range of SLCV only include cucurbitaceous plants (e.g., melon, watermelon, squash and zucchini) it was TYLCV clones infectivity displayed typical symptomatology of this virus in tomato plants yellowing, epinasty and leaf curly (Figure 4). The syndrome development on greenhouse was at 8

days post-inoculation (dpi) and 15 dpi in the growth chamber. On the other hand, SLCV clones did not display any symptom, however when co-inoculated with TYLCV showed more severe symptoms than those observed in single TYLCV infections (Figure 4). Symptoms observed in these latter treatments were small leaves, intervenial chlorosis, strong yellowing, leaf curly and dwarf. The symptoms appearance was similar to those for single TYLCV-infected plants. In all cases, symptomatic plants DNA extractions and PCR analysis confirmed the presence of the respective virus (data not shown).

Discussion

During the years 2005-2006 TYLCV was first detected affecting tomato crops in Sinaloa (Gómez-Jiménez et al, 2006). Thus, it is likely further infections in these crops based on TYLCV wide host range including several weeds common in Sinaloa region and the presence of whiteflies populations (Moriones and Nava-Castillo, 2010). In our work, 72 % of TYLCV incidence was detected, that was not unexpected based on the aforementioned. Sinaloa agroecosystems are conducive for mixed infections and emergence of new begomoviruses (Brown et al. 2002; Rojas and Gilbertson, 2008). Begomoviruses mixed infections commonly reported in pepper and tomato crops in México are between PHYVV and PepGMV (Morales et al. 2005; Morales, 2010). Interestingly, in our study, no mixed infections were detected between TYLCV and PHYVV. On the other hand, seven samples displayed mixed infections between TYLCV and two begomoviruses of SLCV lineage (five with SLCV and two with PepGMV). The high incidence of TYLCV detected, might be explained by several factors involved in the agroecosystems in

Sinaloa state of México. First, the high extension of tomato and “tomatillo” (*Physalis ixocarpa*) crops, as well as the presence of reservoir weeds for TYLCV (Gámez-Jiménez et al. 2009). Moreover, The prevalence in recent years in Sinaloa of biotype B of *B. tabaci*, which is a more efficient TYLCV-transmitter than biotype A (Rojas and Gilbertson, 2008). Another aspect is the presence of biotype Q of *B. tabaci* in Sinaloa, which is efficient in transmitting TYLCV and SLCV, and have tomato and cucurbits as recognized hosts (Brown, 2007a; Martínez-Carrillo and Brown, 2007). A whitefly may acquire and transmit several geminiviruses depending on conditions as high virus concentration in plant hosts and the presence of adequate young tissues for easy acquisition of these viruses (Czosnek, 2007; Brown, 2007b). Once within the plant, TYLCV moves through the plant showing tropism to phloem and cambium (Wege, 2007). Moreover it has been reported that TYLCV V2 protein suppresses RNA-gene silencing system in tomato (Glick et al. 2008). Thus several possibilities are open to explain this phenotype for TYLCV-SLCV tomato infections. On one hand, movement proteins of SLCV could be dysfunctional in tomato, likely because this is a new “host” for this virus and poor co-evolution among them has occurred as suggested in other geminivirus-host interactions (Mendez-Lozano et al. 2003). On the other hand, pathogenicity determinants as suppressor proteins of RNA-silencing system as AC2 of SLCV, might not be functional in hosts as tomato, thus needing some “help” by similar proteins as V2 of TYLCV, then inducing more severe symptoms (Cañizares et al. 2008; Martínez-Turino and Hernández, 2009). This latter aseveration agree with observations reported in other plant-virus systems (Hagen et al. 2008). One or both of the aforementioned possibilities could explain this

synergistic behavior of mixed infections in tomato. More studies are needed to unravel this phenomenon.

References

Argüello-Astorga G, Herrera-Estrella L, Rivera-Bustamante R. (1994a) Experimental and theoretical definition of geminivirus origin of replication. *Plant Mol Biol* 26:553-556.

Argüello-Astorga GR, Guevara-Gonzalez RG, Herrera-Estrella LR, Rivera-Bustamante RF. (1994b) Geminivirus replication origins have a Group-specific organization of iterative elements: a model for replication. *Virology* 203:90-100.

Argüello-Astorga GR, Ruíz-Medrano R. (2001) An iteron related domain is associated to motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. *Arch Virol* 146:1465-1485.

Brown JK. (2007a) The *Bemisia tabaci* complex: Genetic and phenotypic variability drives Begomovirus spread and virus diversification. On line. APS Features. Doi:10.1094/APSnetFeature-2007-017.

Brown JK. (2007b) The *Bemisia tabaci* complex: Genetic and phenotypic variation and relevance to TYLCV-vector interactions. In Czosneck H, ed. Tomato yellow leaf curl virus disease. Pp 25-26. Springer ISBN 978-1-4020-4769-5 (e-book).

Brown JK, Idris AM, Alteri C, Stenger DC. (2002) Emergence of a new cucurbit-infecting begomovirus species capable of forming viable reassortants with related viruses in the *Squash leaf curl virus* cluster. *Phytopathology* 92:734-742.

Cañizares MC, Navas-Castillo J, Moriones E. (2008) Multiple suppressors of RNA silencing encoded by both RNAs of the crinivirus, *Tomato chlorosis virus*. *Virology* 379:168-174.

Cardenas-Conejo Y, Argüello-Astorga G, Poghosyan A, Hernández-González J, Lebsky V, Holguin-Peña J, Medina-Hernández D, Vega-Peña S. (2010) First report of Tomato yellow leaf curl virus co-infecting pepper with *Tomato chino La Paz virus* in Baja California Sur, México. *Plant Disease* 94:1266.

Czosneck H. (2007) Interaction of *Tomato yellow leaf curl virus* with its whitefly vector. In Czosneck H, ed. *Tomato yellow leaf curl virus* disease. Pp 157-170. Springer ISBN 978-1-4020-4769-5 (e-book).

Dellaporta SL, Wood J, Hicks JB. (1983) A plant DNA miniprep: version II. *Plant Mol Biol Rep* 1:19-21.

Gámez-Jiménez C, Romero-Romero JL, Santos-Cervantes ME, Leyva-López NE, Méndez-Lozano J. (2009) Tomatillo (*Physalis ixocarpa*) as a natural new host for Tomato yellow leaf curl virus in Sinaloa, México. *Plant Dis* 93:545.

Glick E, Zrachya A, Levy Y, Mett A, Gidoni D, Belausov E, Citovsky V, Gafni Y. (2008) Interaction with host SGS3 is required for suppression of RNA silencing by *Tomato yellow leaf curl virus* V2 protein. *Proc Natl Acad Sci USA* 105:157-161.

Hagen C, Rojas MR, Kon T, Gilbertson RL. (2008) Recovery from *Cucurbit leaf crumple virus* (family *Geminiviridae*, genus *Begomovirus*) infection is an adaptive antiviral response associated with changes in viral small RNAs. *Phytopathology* 98:1029-1037.

Idris AM, Mills-Lujan K, Martin K, Brown JK. (2008). *Melon chlorotic leaf curl virus*: characterization and differential reassortment with closest relatives reveal adaptive virulence in the Squash leaf curl virus clade and host shifting by the host-restricted *Bean calico mosaic virus*. *J Virol* 82:1959-1967.

Idris AM, Abdel-Salam A, Brown JK. (2006). Introduction of the New world *Squash leaf curl virus* to squash (*Cucurbita pepo*) in Egypt: a potential threat to important food crops. *Plant Dis* 90:1262.

Martínez-Carrillo JL, Brown JK. (2007) First report of the Q biotypes of *Bemisia tabaci* in Southern Sonora, México. *Phytoparasitica* 35:282-284.

Martínez-Turiño S, Hernández C. (2009). Inhibition of RNA silencing by the coat protein of *Pelargonium flower break virus*: distinctions from closely related suppressors. *J Gen Virol* 90:519-525.

Mauricio-Castillo JA, Argüello-Astorga GR, Ambriz-Granados S, Alpuche-Solís AG, Monreal-Vargas CT. (2007). First report of *Tomato golden mottle virus* on *Lycopersicon esculentum* and *Solanum rostratum* in México. *Plant Dis* 91:1513.

Mauricio-Castillo JA, Argüello-Astorga GR, Alpuche-Solís AG, Monreal-Vargas CT, Díaz-Gómez O. (2006). First report of Tomato severe leaf curl virus in México. *Plant Dis* 90:1116.

Mendez-Lozano J, Torres-Pacheco I, Fauquet CM, Rivera-Bustamante RF. (2003) Interactions between geminiviruses in a naturally occurring mixture: *pepper huasteco virus* and *pepper golden mosaic virus*. *Phytopathology* 93:270-277.

Moriones E, Navas-Castillo J. (2010) *Tomato yellow leaf curl* disease epidemics. In: Stanley PA, Naranjo SE. Eds. *Bemisia: Bionomics and Management of a global pest*. Pp. 259-282. Springer Science+Business Media B.V. ISBN 978-90-481-2459-6.

Padidam M, Beachy RN, Fauquet CM. (1995) Classification and identification of geminiviruses using sequence comparisons. *J Gen Virol*. 76:249-263.

Rojas MR, Hagen C, Lucas WJ, Gilbertson RL.(2005) Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. *Annu Rev Phytopathol*. 43:361-94.

Rojas MR, Gilbertson RL. (2008) Emerging plant virus: a diversity of mechanisms and opportunities. In: Roossinck MJ, ed. *Plant virus evolution*. Pp. 27-51. Springer-Verlag, Berlín, Heidelberg.

Sambrook J, Russell D. (2001) *Molecular cloning: a laboratory manual*, 3rd edn. Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press.

Wege C. (2007) Movement and localization of Tomato yellow leaf curl viruses in the infected plant. In: Czosneck (ed.) *Tomato yellow leaf curl virus disease*. Pp 185-206. Springer ISBN 978-1-4020-4769-5 (e-book).

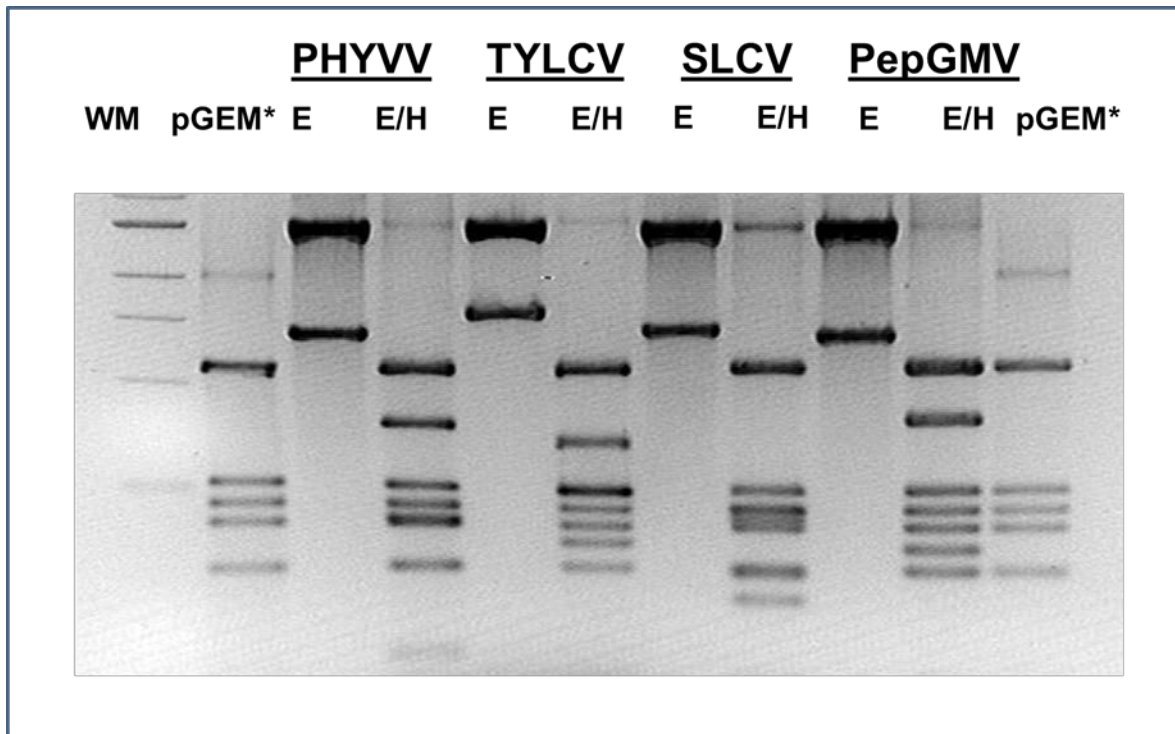


Figure 2. RFLP analysis of repDGRfw-cpYMACrv (“typical” BGVs) and repSL2150fw-cpYMACrv (SLCV clade BGVs) amplicons cloned in pGEM-T Easy vector using *EcoRI* and *HinFI*. **a-** pGEM: empty vector; PHYVV, cloned repDGRfw-cpYMACrv amplicon from a TYLCV and PHYVV-containing sample; TYLCV, cloned repDGRfw-cpYMACrv amplicon from a SLCV and TYLCV-containing sample; SLCV, cloned repSL2150fw-cpYMACrv amplicon from a SLCV and TYLCV-containing sample; PepGMV, cloned repSL2150fw-cpYMACrv amplicon from a PepGMV and TYLCV-containing sample

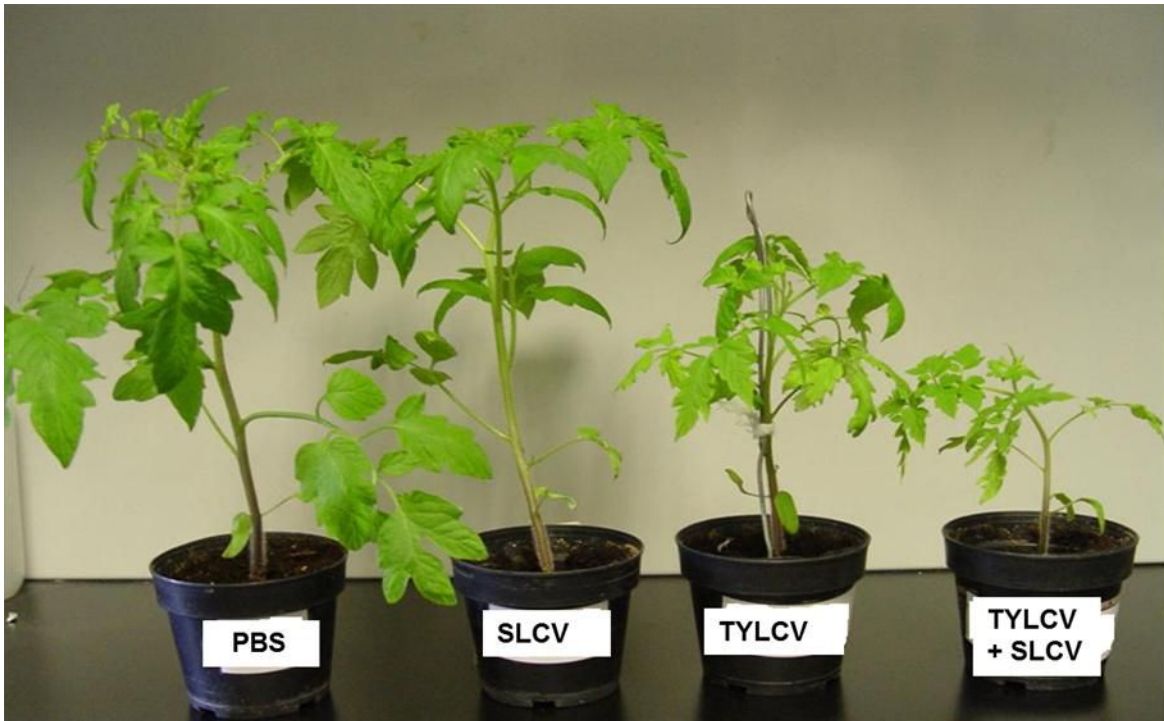


Figure 3. Symptoms induced by infective clones of TYLCV and SLCV isolated from tomato plants. All plants were inoculated as described in Materials and Methods.