



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

POSGRADO EN CIENCIAS EN BIOLOGÍA MOLECULAR

**Reconstrucción de la evolución genómica de
los begomovirus por medio de un enfoque
multidisciplinario integrado**

Tesis que presenta
Yair Cárdenas Conejo

Para obtener el grado de
Doctor en Ciencias en Biología Molecular

Directores de Tesis:

Dr. Gerardo Rafael Argüello Astorga
Dra. Lina Raquel Riego Ruiz

San Luis Potosí, S.L.P., 10 de Agosto, 2012



Constancia de aprobación de la tesis

La tesis “**Reconstrucción de la evolución genómica de los begomovirus por medio de un enfoque multidisciplinario integrado**” presentada para obtener el Grado de Doctor en Ciencias en Biología Molecular fue elaborada por **Yair Cárdenas Conejo** y aprobada el **10 de agosto de 2012** 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)

Dra. Lina Raquel Riego Ruiz
(Director de la tesis)

Dra. Irene Beatriz Castaño Navarro
(Miembro del Comité Tutorial)

Dra. Ruth Elena Soria Guerra
(Miembro del Comité Tutorial)

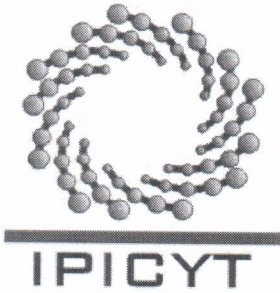


Créditos Institucionales

Esta tesis fue elaborada en el Laboratorio de Biología Molecular de plantas de la División de Biología Molecular del Instituto Potosino de Investigación Científica y Tecnológica, A.C., bajo la codirección de los doctores Gerardo Rafael Argüello Astorga y Lina Raquel Riego Ruiz.

Durante la realización del trabajo el autor recibió una beca académica del Consejo Nacional de Ciencia y Tecnología (No. de registro 209543) y del Instituto Potosino de Investigación Científica y Tecnológica, A. C.

El proyecto de investigación descrito en esta Tesis fue financiado con recursos otorgados al Dr. Gerardo Argüello Astorga por el CONACYT (Proyecto SEP 84004)



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 047 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 10 días del mes de agosto del año 2012, se reunió a las 12:00 horas en las instalaciones del Instituto Potosino de Investigación Científica y Tecnológica, A.C., el Jurado integrado por:

Dra. Irene Beatriz Castaño Navarro	Presidenta	IPICYT
Dra. Ruth Elena Soria Guerra	Secretario	UASLP
Dr. Gerardo Rafael Argüello Astorga	Sinodal	IPICYT
Dra. Lina Raquel Riego Ruiz	Sinodal	IPICYT

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

DOCTOR EN CIENCIAS EN BIOLOGÍA MOLECULAR

sustentó el C.

Yair Cárdenas Conejo

sobre la Tesis intitulada:

Reconstrucción de la evolución genómica de los begomovirus por medio de un enfoque multidisciplinario integrado

que se desarrolló bajo la dirección de


Dra. Lina Raquel Riego Ruiz
Dr. Gerardo Rafael Argüello Astorga

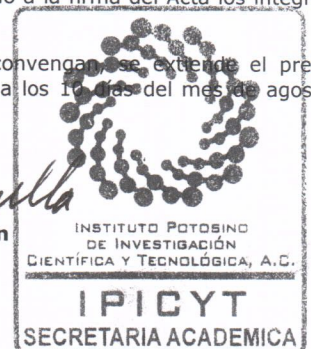
El Jurado, después de deliberar, determinó

APROBARLO

Dándose por terminado el acto a las 13:40 horas, procediendo a la firma del Acta los integrantes del Jurado. Dando fe el Secretario Académico del Instituto.

A petición del interesado y para los fines que al mismo convengan, se extiende el presente documento en la ciudad de San Luis Potosí, S.L.P., México, a los 10 días del mes de agosto de 2012.


Dr. Marcial Bonilla Marín
Secretario Académico




Mtra. Ivonne Lizette Cuevas Vélez
Jefa del Departamento de Posgrado

DEDICATORIAS

La finalización de esta etapa de estudios, resumida en las páginas de este trabajo de investigación, se la dedico a mi familia que está conformada por mis queridos padres Ezequiel Cárdenas Mendoza y Ma. Elizabeth Conejo Jaimes y hermanos Nadia Itzel Cárdenas Conejo, Alán Cárdenas Conejo y Zugey Elizabeth Cárdenas Conejo, todos ellos me han ayudado a ser la persona feliz y plena que soy en estos momentos.

Este trabajo también se lo dedico a mis directores de tesis Gerardo Rafael Argüello Astorga y Lina Raquel Riego Ruiz pues juntos logramos desenmarañar algunos misterios de la evolución de los begomovirus.

Agradecimientos

A mis padres Ezequiel Cárdenas Mendoza y Ma. Elizabeth Conejo Jaimes por enseñarme el significado de la vida y por alentarme a terminar esta etapa de mi vida

Al Dr. Gerardo por la amistad que me ha brindado y por la confianza que deposito en mi persona para realizar esta investigación. También le agradezco que me enseñara la forma de razonar ante los problemas que esta investigación arrojó, un razonamiento al más puro estilo detectivesco de Chevalier Auguste Dupin.

A la Dra. Lina por regalarme su amistad y por estar al pendiente y dispuesta a ayudar a resolver los problemas que surgieron en cada paso de la investigación.

A mis compañeros y amigos del laboratorio de Biología Molecular de plantas que me ayudaron en el trabajo experimental, en especial agradezco a mis compañeros Jedi Jorge Armando, Mariana Cantú, Bernardo Bañuelos, Gregorio Josefát, Omar Patiño, Edith Uresti e Iliana Torres.

A Salvador por su amistad y por enseñarme las técnicas de Biología Molecular que conozco.

A los amigos que me acompañaron a lo largo de estos seis años por brindarme su amistad y cariño, Andrómeda Liñan, Jorge Armando, Mariana Cantú, Bernardo Bañuelos, Gregorio Josefát, Omar Patiño, Aída Araceli, Jorge Alberto, Javier Montalvo, Luzmila, LULA, Karla Sánchez, Edith Uresti, Iliana Torres, Raúl Loera, Alma Jasso, Marcela Briones, José Ángel Huerta, Jaqueline, Verónica Martínez, Verónica Márquez, Mayra, Liz, etc...

A la Sra. Rosy por estar al pendiente de mi espacio de trabajo y por la amistad que me dio desde el momento en que llegue al laboratorio.

Índice

Constancia de aprobación de la tesis	ii
Créditos Institucionales	iii
Copia del Acta de Examen	iv
DEDICATORIAS	v
Agradecimientos	vi
Resumen	x
ABSTRACT	xi
Introducción	1
Virus	1
Virus de plantas	3
Familia Geminiviridae	3
<i>Mastrevirus</i>	5
<i>Curtovirus</i>	6
<i>Topocuvirus</i>	7
<i>Begomovirus</i>	9
<i>Begomovirus Bipartitas</i>	9
<i>Begomovirus monopartitas</i>	10
<i>Begomovirus del Nuevo y del Viejo Mundo</i>	11
<i>Proteínas de Begomovirus</i>	11
<i>Satélites</i>	20
<i>Alfa-satélites</i>	20
<i>Beta-satélites</i>	21
<i>Otros satélites: ToLCV-Sat</i>	21
Origen de los geminivirus	22
Planteamiento de los problemas que se abordan en el estudio.	25
Objetivo del Estudio	26
RESULTADOS-PARTE I	27
Centro de Origen de los begomovirus y organización genómica del ancestro común.	27
I.1- Filogenia de los componentes genómicos de begomovirus	27
I.2 - Identificación de marcadores biogeográficos y “firmas” específicas de linaje en las proteínas BV1 y BC1.	33
I.2.1 - Análisis de las proteínas BV1	33
I.2.2-Análisis de las proteínas BC1	35
I.2.3 Análisis de las proteínas CP	36
I.3- Análisis comparado de las regiones no codificantes	39
1.3.1. Análisis comparativo del promotor <i>BV1</i>	40
1.3.2 Análisis comparativo del promotor <i>BC1</i>	42

I.4 – Identificación de la India como Centro de Origen de los begomovirus _____	46
I.5 – Organización genómica del Ancestro Común de los BGVs _____	50
Parte 2 _____	52
Origen de los begomovirus del Nuevo Mundo _____	52
Rutas potenciales de migración al NM _____	54
La ruta de migración “Africana” _____	55
La ruta de migración de Asia a América por el estrecho de Bering _____	56
Conclusión, centro de origen de los begomovirus y migración al NM _____	59
Parte 3 _____	61
Evolución del segundo componente de los Begomovirus _____	61
<i>Las proteínas BV1 y CP son homólogas _____</i>	61
<i>Los promotores de BV1 y CP son homólogos _____</i>	67
<i>Origen del Gen BV1 _____</i>	71
<i>La proteína BC1 se originó de los genes sobrepuestos AC1, AC2 y AC3 _____</i>	71
<i>Análisis comparativo entre AC3 y la C-terminal de BC1 _____</i>	73
<i>Análisis comparativo entre AC2 y BC1 _____</i>	79
<i>Análisis comparativo entre AC1 y la región N-terminal de BC1 _____</i>	81
<i>Conclusión, origen de la proteína BC1 _____</i>	84
<i>El promotor de BC1 se originó del gen AC1 _____</i>	88
<i>Conclusión, origen del gen BC1 _____</i>	90
Conclusión, el componente B de begomovirus se origino evolutivamente de un componente A _____	93
Métodos _____	95
Referencias _____	108

Lista de figuras

<i>Figura 1. Estructura de la cápside de los geminivirus. _____</i>	4
<i>Figura 2. Organización genómica de los géneros de la familia Geminiviridae. _____</i>	8
<i>Figura 3. Modelo del movimiento de begomovirus. _____</i>	19
Figura 4. Reconstrucción filogenética de los componentes A y B de begomovirus. _____	30
Figura 5 Reconstrucción filogenética de las proteínas BV1 y BC1 de begomovirus. _____	32
<i>Figura 6. Motivo 130-145 de la proteína BV1. _____</i>	34
<i>Figura 7. Motivos específicos de linaje de la proteína BC1. _____</i>	36
<i>Figura 8. Análisis comparativo de la proteína CP _____</i>	38
<i>Figura 9. Arreglo modular conservado de Legumovirus _____</i>	41
<i>Figura 10. Ubicación de la CPS en el promotor de BC1 _____</i>	43
Figura 11. Localización de la CPS en el componente A. _____	43
<i>Figura 12. Predicción de la estructura secundaria de la CPS _____</i>	45
<i>Figura 13. Estructura del promotor ancestral de CP. _____</i>	49

Lista de tablas

<i>Tabla 1 Lista de las proteínas de begomovirus</i>	19
<i>Tabla 2 Secuencias utilizadas en la reconstrucción filogenética</i>	100
<i>Tabla 3 Secuencias de aminoácidos y nucleótidos utilizadas en las figuras presentadas</i>	103

Resumen

Reconstrucción de la evolución genómica de los begomovirus por medio de un enfoque multidisciplinario integrado.

Los geminivirus son patógenos de plantas superiores caracterizados por una cápside geminada y un genoma compuesto por una o dos moléculas circulares de ADN de cadena sencilla (ADNcs). El género *Begomovirus*, con más de 200 especies, es el subgrupo de la familia *Geminiviridae* más diverso y de distribución mundial más amplia. Este género incluye a los únicos geminivirus que poseen genomas compuestos por dos moléculas de ADNcs. El ADN-A codifica proteínas necesarias para la replicación viral, el control de la expresión de genes tardíos, la formación de la cápside y la supresión de respuestas antivirales del huésped. El DNA-B, por otra parte, codifica dos proteínas, BV1 y BC1, que están involucradas en el movimiento intracelular e intercelular del virus. Diversas líneas de evidencia sugieren que los begomovirus (BGVs) evolucionaron hace 130-110 millones de años, pero quedan por resolver todavía varias cuestiones fundamentales de su historia evolutiva, incluyendo la ubicación de su centro de origen, su migración al continente americano, y la importante y controversial cuestión relativa al origen del segundo componente genómico. Mediante el empleo de una amplia gama de programas para la inferencia de filogenias a partir de las secuencias de ADN y proteínas, el análisis filogenético-estructural de secuencias no-codificantes, la identificación de marcadores moleculares biogeográficos, la información proporcionada por los fósiles virales, y los datos paleobiogeográficos del vector, logramos establecer que el centro de origen de los BGVs se localiza en la India. El análisis condujo también a la conclusión de que los BGVs de América se derivaron de BGVs originarios del subcontinente Indio, y postulamos la hipótesis de que uno o más linajes de la India arribaron a Sudamérica en la Era Cenozoica, a través del continente Antártico. Finalmente, utilizando diversos métodos para reconstruir secuencias ancestrales y trazar homologías remotas, logramos establecer que el ADN-B se derivó evolutivamente de un ADN-A ancestral, a través de un proceso que involucró divergencia y especialización génica en el caso del gen *BV1*, y la generación por un mecanismo de “sobre-inscripción” (“overprinting”) de un nuevo gen, *BC1*, a partir de tres genes ancestrales (*AC1*, *AC2* y *AC3*) con funciones no relacionadas al movimiento viral. Este último proceso representa un caso inédito de evolución molecular, esto es, la emergencia de un gen con una nueva función a partir de las secuencias de tres genes traslapados con funciones diferentes.

Palabras clave: begomovirus, evolución viral, evolución de genes, proteínas de movimiento, overprinting.

ABSTRACT

Geminiviruses are important plant pathogens characterized by a geminate capsid and a genome composed by one or two circular molecules of single-stranded DNA (ssDNA). The genus *Begomovirus*, with over 200 species, is the subgroup of the family *Geminiviridae* that is more diversified and widely distributed in the world. Begomoviruses are the only geminiviruses possessing genomes composed by two DNAs molecules (DNA-A and DNA-B). The DNA-A encodes proteins necessary for virus replication, control of gene expression, suppression of gene silencing, and encapsidation. The DNA-B, on the other hand, encodes two proteins, BV1 and BC1, which are involved in the intracellular and intercellular movement of the virus. Several lines of evidence suggest that begomoviruses (BGVs) evolved 130-110 million years ago, but still remain to be resolved several key issues of their evolutionary history, including the location of its center of origin, its migration to the Americas, and the enigmatic issue of the second genomic component evolutionary origin. Employing a wide range of bioinformatics software for inferring phylogenies from DNA and proteins sequences, the phylogenetic-structural analysis of non-coding sequences, the identification of biogeographic molecular markers, the information from viral fossils and the paleobiogeographic data on the whitefly vector, we concluded that the BGVs center of origin is located in India. The analysis also led to the conclusion that American BGVs were derived from Indian subcontinent BGVs, and we postulate the hypothesis that one or more lineages of BGVs arrived to South America in the Cenozoic through the Antarctic continent. Finally, using various methods for reconstructing ancestral sequences and trace remote homologies between genes, we established that DNA-B was derived from an evolutionarily ancient DNA-A, through a process involving gene divergence and specialization in the case of the *BV1* gene, and generation of a new gene, *BC1*, by overprinting from three ancestral genes (*AC1*, *AC2* and *AC3*) with functions unrelated to viral movement. This latter process represents an unprecedented case of molecular evolution, ie, the emergence of a gene with a new function from the sequences of three overlapping genes with different functions.

Key words: begomovirus, geminivirus, viral evolution, gene evolution, overprinting

Introducción

Virus

El estudio de los virus comenzó en 1898 cuando Martinus W. Beijerinck descubrió que la causa de la enfermedad del moteado de las hojas de tabaco no eran bacterias sino un agente infeccioso al que denominó "contagium vivum fluidum". Beijerinck también descubrió que este agente infeccioso tenía la capacidad de atravesar filtros de porcelana, por lo que debía de ser más pequeño que una bacteria, además de que el patógeno no se podía cultivar en un medio sintético, sino solamente en el tejido en desarrollo de la planta del tabaco (Beijerinck 1898). Posteriormente el "contagium vivum fluidum" fue denominado *Virus del mosaico del tabaco* (TMV), el primer virus caracterizado.

Los virus (del latín *virus*, toxina o veneno) son elementos genéticos móviles (ADN o ARN) envueltos en una estructura proteica que los protege llamada cápside, capaces de moverse de una célula a otra. La partícula formada por la cápside y el material genético tiene un tamaño sub-microscópico, de alrededor de 15 a 300 nanómetros, aunque algunos virus alcanzan los 750 nanómetros de diámetro como es el caso de *Megavirus chilensis* (Arslan et al. 2011). El material genético de los virus lleva codificada la información para producir las proteínas necesarias para su replicación, la cual se lleva a cabo en una célula huésped que le provee la maquinaria bioquímica para realizar esta tarea, ya que los virus son incapaces de ser funcionalmente activos fuera de la célula huésped. Existe una gran variedad de virus que se distinguen por el tipo de ácido nucleico que constituye su genoma, la estructura de su envoltura proteica, el modo de entrar a la célula, y la forma en que se replican estando dentro de ella.

La mayoría de los virus se consideran patógenos ya que cuando invaden las células vivas y las usan para multiplicarse provocan daños en la célula huésped, como alteraciones morfológicas y funcionales, o lisis celular, que provoca que el organismo huésped manifieste síntomas e incluso muera. Entre las enfermedades virales más importantes que afectan al ser humano figuran el síndrome de la inmunodeficiencia adquirida (SIDA), la hepatitis B, la influenza, el síndrome respiratorio agudo severo (SARS por sus siglas en inglés), viruela, varicela y poliomielitis. Los virus pueden infectar muchos tipos de organismos como bacterias, algas, protozoarios, hongos, invertebrados, vertebrados y plantas. La mayoría de los virus se restringen a un huésped particular, pero existen algunos que tienen la capacidad de replicarse en diferentes organismos. Por ejemplo, los geminivirus pueden replicarse tanto en células vegetales como en la bacteria *Agrobacterium tumefaciens* (Rigden et al. 1996) y la levadura *Saccharomyces cerevisiae* (Alves-Rodrigues et al. 2006). Otros virus como *Tomato spotted wilt virus*, *Maize rayado fino virus* y *Tomato yellow leaf curl virus* se pueden replicar tanto en el insecto vector como en la célula huésped (Medeiros et al. 2004; Kitajima & Gámez 1983; Rubinstein & Czosnek 1997). Recientemente se reportó la presencia de un virus de plantas, *Pepper mild mottle virus* (PMMoV) en altas concentraciones (hasta 10^9 viriones por gramo) en heces fecales humanas (Zhang et al. 2006), e incluso se reportó que es capaz de inducir anticuerpos para PMMoV lo cual indica que el virus no es un componente neutral de la flora intestinal (Colson et al. 2010); sin embargo, no existe evidencia de que este virus sea capaz de replicarse en células humanas.

Virus de plantas

Los virus de plantas causan varias clases de enfermedades, las cuales por lo general no conducen a la muerte de la planta. Los síntomas más comunes son manchas anulares, desarrollo de patrones de mosaicos, amarillamiento y deformaciones de la hoja. Las plantas completas pueden llegar a sufrir trastornos del crecimiento y mostrar enanismo. Algunos de los virus de plantas pueden infectar diferentes especies. Por ejemplo, el curtovirus *Beet curly top virus* infecta betabel, amaranto, chile, tomate, frijol, alfalfa, tabaco, y muchas otras más (Thornberry 1966; Wallace & Murphy 1938). El material genético de los virus de plantas puede ser de ADN de doble cadena, ARN de doble cadena, ADN de cadena sencilla y ARN de cadena sencilla. La mayoría de los virus de plantas tienen genomas de ARN de cadena sencilla o doble, son pocas las familias de virus de ADN, una de estas es la familia *Geminiviridae*, cuyos miembros poseen un genoma de ADN circular de cadena sencilla.

Familia Geminiviridae

Los geminivirus son virus de plantas que infectan plantas monocotiledóneas y dicotiledóneas, muchas de ellas de valor agrícola, como el maíz, el frijol, el chile, el tomate, etc. Los miembros de la familia *Geminiviridae* poseen un genoma compuesto por una o dos moléculas circulares de ADN de cadena sencilla (ADNcs) de un tamaño que varía entre 2.5 y 3.0 kb. La característica distintiva de esta familia es la cápside que protege su genoma, la cual es una cápside con un diámetro de ~18nm y una longitud de ~30nm, que presenta la estructura de dos icosahedros cuasi-isométricos fusionados por una de sus caras, lo que le da la apariencia de una partícula “geminada” en las

micrografías electrónicas (Zhang et al. 2001). Esta morfología característica del virión es la que dio su nombre a esta familia viral (Fig. 1). La cápside geminada consta de un total de 110 subunidades de la proteína CP, organizadas como 22 pentámeros (Zhang et al. 2001). Esta proteína está conservada y su capacidad de formar los viriones gemelos es única (Harrison 1985), pero ha divergido en términos de la especificidad por el insecto vector que lo transmite (Briddon et al. 1990). Así como la proteína CP está presente en todos los geminivirus, la familia posee una proteína esencial para la replicación (Rep), que es conservada en secuencia, posición y función (Gutierrez 1999; Hanley-Bowdoin et al. 1999). La proteína Rep reconoce y se une al origen de replicación (*Ori*) mediante unas secuencias repetidas llamadas iterones (Argüello-Astorga et al. 1994). Dentro del *Ori* también se encuentra una secuencia con el potencial de formar una estructura de tallo-asa (“horquilla”) que contiene un nona-nucleótido (TAATATT*AC) donde Rep introduce un corte (“Nick”) en una de las cadenas, que es un evento necesario para que se inicie el proceso de replicación por círculo rodante (RCR) del DNA viral (Hanley-Bowdoin et al. 1999).

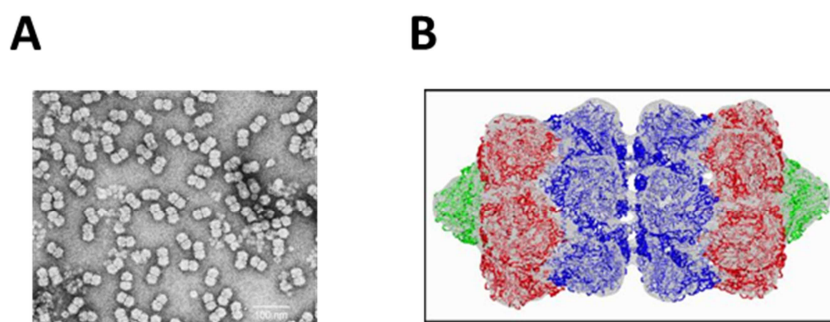


Figura 1. Estructura de la cápside de los geminivirus.

- a) Micrografía electrónica de geminivirus, obtenida de la galería de figuras de la web del ICTVdB revisada en junio del 2012.
- b) Modelo de la estructura de la cápside de geminivirus (Zhang et al. 2001)

La familia *Geminiviridae* se divide en cuatro géneros: *Mastrevirus*, *Curtovirus*, *Topocuvirus* y *Begomovirus*, de acuerdo a su organización genómica, rango de huéspedes, e insecto vector que los transmite. Las características principales de los géneros son mencionadas a continuación y la organización genómica de sus miembros se muestran en la figura 2.

Mastrevirus

Los mastrevirus son transmitidos por diversas especies de chicharritas (Hemiptera, *Cicadellidae*) e infectan plantas monocotiledóneas o dicotiledóneas, a diferencia de otros geminivirus que se restringen a las dicotiledóneas. El genoma de estos virus es monopartita, y contiene cuatro ORFs: dos en el sentido del virión (*V1* y *V2*), y dos en el sentido complementario (*C1* y *C2*). *V1* codifica la proteína de movimiento (MP) y *V2* codifica la proteína de la cápside, en tanto que el ORF *C1* codifica la proteína asociada a la replicación llamada Rep A, y los ORFs *C1* y *C2* codifican conjuntamente a la proteína Rep, la cual es traducida a partir de un transcrito primario *C1-C2* que experimenta la delección de un pequeño intrón (“splicing”) localizado entre ambos ORFs. A diferencia de otros geminivirus, los miembros de este género tienen dos regiones intergénicas (RI) bien definidas: una RI “corta” y otra RI “larga” (Fig. 2D). Dentro de la RI larga se encuentra el *Ori* y dos promotores divergentes que controlan la expresión de los genes que se transcriben en sentidos opuestos. Los mastrevirus también se distinguen de los otros subgrupos geminivirales por que en el virión se empaqueta un oligonucleótido de ~80 nt que es complementaria a una secuencia presente en la RI

corta; este oligonucleótido funciona como un iniciador o “primer” de la replicación de la cadena complementaria (Donson et al. 1984). Otra particularidad distintiva de los mastrevirus es la presencia de un intrón en el gen *C1/C2*, que genera por “splicing” diferencial 2 proteínas (Gutierrez 1999): 1) la proteína Rep completa, involucrada en el proceso de replicación por CR, y que también participa en la regulación de la expresión de los genes tardíos, y 2) la proteína Rep-A, que se une a la proteína relacionada al retinoblastoma (pRBR) y está involucrada en la interferencia del ciclo celular del huésped (Liu, Saunders, et al. 1999; Gutierrez et al. 2004). La proteína CP de los mastrevirus es una proteína multifuncional, tiene la capacidad de formar la cápside y participar en el movimiento del virus ya que tiene la función de lanzadera (importar y exportar el ADN al núcleo), y participa en el movimiento sistémico en combinación con la proteína MP (V1) (Liu, Boulton, et al. 1999; Liu et al. 2001).

Curtovirus

Los curtovirus son virus que se han aislado tanto en el Viejo Mundo (VM) como en el Nuevo Mundo (NM); son transmitidos por chicharritas del género *Circulifer*. Los curtovirus sólo infectan plantas dicotiledóneas pero tienen la capacidad de infectar un amplio rango de huéspedes. La comparación por pares de los genomas indica que los curtovirus posiblemente se originaron de un evento de recombinación entre un mastrevirus arcaico y un virus parecido a los begomovirus (Padidam et al. 1995; Stanley et al. 1986; Palmer & Rybicki 1998). El genoma monopartita de los curtovirus tiene una organización diferente al de los mastrevirus, y contiene una IR y siete genes que codifican igual número de proteínas. Tres de sus genes se encuentran en el sentido

del virión (*MP* ó *V3*; *CP* ó *V1*; y el gen *V2*) y cuatro en el sentido complementario (*Rep* ó *C1*, *C2*, *C3* ó *REn*, y *C4*) (Fig. 2C). Los genes están separados por una RI en la cual se encuentran promotores divergentes y el origen de replicación. La proteína Rep está implicada en la replicación del virus, mientras que la proteína C3 o REn tiene la función de potenciar la replicación, ya que su mutación reduce los niveles de DNA viral presentes en las plantas infectadas (Stanley et al. 1992). La proteína C2 es un determinante de patogenicidad, ya que mutantes *C2*⁻ de *Spinach curly top virus* presentan una reducción en la infectividad y la severidad de los síntomas (Baliji et al. 2007), sin embargo, las mutantes *C2* de *Beet curly top virus* presentan el mismo fenotipo que los virus silvestres (Hormuzdi & Bisaro 1995; Stanley et al. 1992). El movimiento de los curtovirus en las plantas infectadas se lleva a cabo por un mecanismo cooperativo de las proteínas CP y MP (Hormuzdi & Bisaro 1993). La función específica de las proteínas CP y MP en el movimiento de los curtovirus no se ha establecido, pero probablemente funcionen de forma similar a las proteínas homólogas de los mastrevirus. La función del producto del gen *V2* no es muy clara, pero se ha propuesto que podría estar involucrado en la regulación de los niveles relativos de ANDcs y ADNcd virales (Hormuzdi & Bisaro 1995).

Topocuvirus

El género *Topocuvirus* está integrado únicamente por el virus monopartita *Tomato pseudo-curly top virus* (TPCTV), el cual es transmitido a plantas dicotiledóneas por la chicharrita saltadora *Micrutalis malleifera* (Hemiptera, *Membracidae*). El genoma de TPCTV incluye seis genes, 2 en sentido del virión (*CP* y *V2*) y cuatro en el sentido

complementario (*C1*, *C2*, *C3* y *C4*), los cuales están divididos por una RI en donde se encuentra el origen de replicación. El análisis genómico de este virus reveló que TPCTV es producto de un evento de recombinación, en donde el módulo de replicación (genes *C1*, *C2*, *C3*) y de patogénesis (*C4*) fue donado probablemente por un curtovirus del NM, mientras que el módulo implicado en el movimiento (genes *CP* y *V2*) no muestra semejanza significativa con el de virus de otros géneros, y corresponde por tanto a una estirpe geminiviral diferente (Bridson et al. 1996).

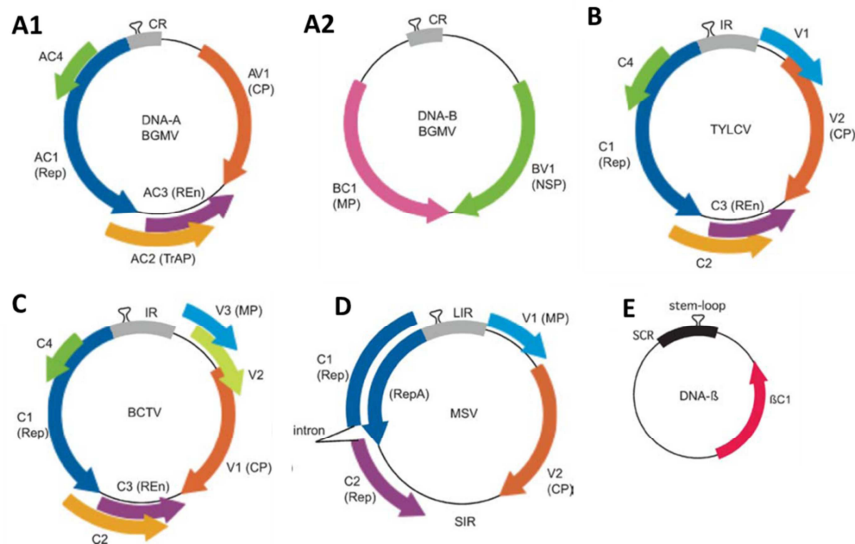


Figura 2. Organización genómica de los géneros de la familia *Geminiviridae*.

A. Genoma de los Begomovirus bipartitas del NM, A1) componente A y A2) componente B.

B. Genoma monopartita de begomovirus del VM (TYLCV)

C. Genoma monopartita del género *Curtovirus* (BCTV)

D. Genoma monopartita del género *Mastrevirus* (MSV)

E. Organización genómica del satélite β

Figura modificada de (Rojas et al. 2005).

Begomovirus

El género *Begomovirus* es el más diversificado y de distribución geográfica más amplia, ya que se encuentran en todas las regiones tropicales y subtropicales del mundo. Los miembros de este género infectan un amplio número de huéspedes, y son transmitidos por una sola especie de insecto vector, la llamada mosquita blanca (*Bemisia tabaci*) (Brown & Bird 1992). Una característica única de este género es que incluye virus monopartitas y virus con genomas compuestos por dos moléculas de ADN, denominadas componente A ó ADN-A y componente B ó ADN-B.

Begomovirus Bipartitas

Los componentes genómicos de los BGVs bipartitas se han especializado para cumplir diferentes funciones durante la infección. El ADN-A conlleva las funciones de replicación viral y transcripción de los genes mientras que el ADN-B se ha especializado en el movimiento local y sistémico del virus. Así, ambos componentes genómicos son esenciales para establecer una infección exitosa. La comparación de las secuencias de nucleótidos entre el ADN-A y el ADN-B indica que estos componentes tienen baja identidad, excepto por un segmento conservado de ~200-nt denominado región común (RC) (Fig. 2A). La RC se encuentra en la RI de ambos componentes (Harrison 1985). La RC alberga el origen de replicación y las secuencias necesarias para la transcripción de los genes divergentes en ambos componentes genómicos (Hanley-Bowdoin et al. 1999). El origen de replicación de los componentes genómicos de una misma especie presenta elementos cortos repetidos, llamados iterones, que constituyen los sitios de unión específica de la proteína Rep cognada. La secuencia nucleotídica de los iterones

suele ser distinta entre especies diferentes, lo que explica por qué el ADN-A de una especie begomoviral no puede replicar al componente B de otra, ya que el arreglo y secuencia de los iterones es diferente y existe incompatibilidad con la proteína Rep heteróloga (Gregorio-Jorge et al. 2010).

El componente A contiene 5 genes que codifican proteínas involucradas en la replicación, transcripción y encapsidación; una proteína en el sentido del virión (AV1 ó CP) y cuatro en el sentido complementario (AC1 ó Rep; AC2 ó TrAP; AC3 ó REn; AC4) divididos por la RI (Fig. 2 A1). El componente B tiene dos genes que codifican para las proteínas involucradas en el movimiento viral y en el desarrollo de síntomas en la planta, uno de ellos en el sentido del virión (*BV1*) y el otro gen en el sentido complementario (*BC1*) (Fig. 2 A2). Adicionalmente el componente A de los BGVs del VM tiene un gen *AV2* en el sentido del virión, el cual codifica para una proteína del movimiento. Sin embargo este gen *AV2* no está presente en los BGVs *Corchorus yellow vein virus* (CoYVV) y *Corchorus golden mosaic virus* (CoGMV), de Vietnam (Ha et al. 2006; Ha et al. 2008), ni en *Soybean chlorotic blotch virus* (SbCBV) de Nigeria (Alabi et al. 2010).

Begomovirus monopartitas

A diferencia de los BGVs bipartitas los cuales se distribuyen tanto en el NM como en el VM, la distribución geográfica de los BGVs monopartitas se restringe al VM. El genoma de estos últimos tiene un arreglo genómico similar al del ADN-A de los virus bipartitas. Todos los virus monopartitas codifican seis proteínas, con dos genes en el sentido del

virión (*V1* ó *CP*; *V2* ó *MP*) y cuatro genes en el sentido complementario (*C1* ó *Rep*; *C2* ó *TrAP*; *C3* ó *REn*; *C4*) divididos por una RI, que contiene el *Ori* y dos promotores divergentes (Fig. 2B). La diferencia entre el genoma monopartita y el genoma bipartita, es que el primero codifica dos proteínas involucradas en el movimiento, esto es, *V2* y *CP*, mientras que en el genoma bipartita el ADN-A no codifica una proteína de movimiento especial, aunque en algunos virus bipartitas la proteína *CP* actúa como proteína lanzadera núcleo-citoplasma.

Begomovirus del Nuevo y del Viejo Mundo

Una serie de características genómicas hace posible dividir los BGVs en virus del NM y VM, estas son: 1) la presencia del gen *AV2* exclusiva de los virus del VM (Rybicki 1994); 2) todos los virus del NM son bipartitas, mientras que en el VM los hay tanto monopartitas como bipartitas; 3) la proteína *CP* de los virus del NM tiene el motivo PWRsMaGT en el N-terminal, el cual está ausente en los virus del VM (Harrison et al. 2002); 4) la mayoría de los virus del VM tienen dos iterones río arriba de la caja TATA del gen *AC1* y un iteron complementario río abajo, este último está ausente en los virus del NM (Argüello-Astorga et al. 1994).

Proteínas de Begomovirus

La mayoría de las proteínas de los BGVs son multifuncionales y la acción conjunta de ellas determina la infección sistémica de la planta huésped. A continuación se describen

las funciones más importantes de cada una de ellas. En la Tabla 1 se resumen las funciones más relevantes de cada proteína begomoviral.

Proteína AV1/cápside

La proteína CP de los BGVs, al igual que la de los otros géneros, es una proteína multifuncional que tiene la capacidad de formar la cápside y de involucrarse en funciones del movimiento viral. En los BGVs monopartitas esta proteína es imprescindible para una infección exitosa (Rigden et al. 1993; Wartig et al. 1997; Rojas et al. 2005). La proteína CP de los virus monopartitas, al igual que la proteína BV1, tiene señales de localización nuclear (SLN), y la capacidad de unirse a importinas α del núcleo, lo que permite que la proteína CP realice la función de exportar e importar el ADN viral al núcleo (Kass et al. 2006; Gafni 2002; Kunik et al. 1998). En los BGVs con genoma bipartita la proteína CP no es esencial para producir una infección sistémica, contrario a lo que se observa en todos los geminivirus monopartitas. BGV bipartitas con mutaciones deletéreas en el gen *CP* son capaces de infectar a su huésped de la misma forma que el virus silvestre, aunque en algunos casos los síntomas se atenúan y su aparición se retrasa significativamente, dependiendo del virus en particular y de la planta huésped (Ingham, Pascal, et al. 1995; Pooma et al. 1996; Qin et al. 1998; Gardiner et al. 1988). Sin embargo, la proteína CP tiene la capacidad de enmascarar el fenotipo de virus a los cuales se les ha mutado la proteína BV1 (Sanderfoot et al. 1996; Qin et al. 1998), lo que sugiere que la CP podría estar complementando la función de esta última en el movimiento del ADN viral hacia dentro y fuera del núcleo, tal como lo hace la proteína CP de los virus monopartitas.

Proteína AV2/pre-cápside

La proteína pre-cápside (PPC) está presente sólo en los virus que son originarios del Viejo Mundo. Se ha demostrado que ésta proteína participa en el movimiento de los begomovirus, específicamente en exportar el ADN viral del núcleo a los plasmodesmos (Rojas et al. 2005; Poornima Priyadarshini et al. 2011). Al fusionar a la PPC de TYLCV con GFP se observó que ésta se localizaba en el retículo endoplásmico y en la periferia de la célula (Rojas et al. 2001). Esta evidencia sugiere que la PPC participa en el movimiento célula-célula. La proteína PPC también se ha descrito como un determinante de patogénesis, tal vez regulando los niveles de ADN de cadena sencilla y los de cadena doble (Wartig et al. 1997).

Proteína AC3/REn

La proteína REn de los BGVs tiene la misma función que la proteína C3 de los curtovirus, esto es, potenciar la replicación viral, ya que se ha visto que mutantes de esta proteína reducen la replicación del ADN viral y la severidad de los síntomas en plantas infectadas (Hormuzdi & Bisaro 1995; Morris et al. 1991; G. Sunter et al. 1990). La proteína REn es capaz de homo-oligomerización, y de interactuar con la proteína Rep y proteínas de la planta como al antígeno de proliferación nuclear (PCNA), a proteínas de la familia NAC y a la proteína homóloga de retinoblastoma (Selth et al. 2005; Castillo et al. 2003; Hanley-Bowdoin et al. 2004; Settlage et al. 1996). Mutantes en la proteína REn afectadas en su capacidad de oligomerización o de interacción con

las proteínas PCNA y Rep reducen o inactivan la actividad de potenciar la replicación por parte de la proteína REn, en tanto que la interacción con la proteína relacionada a Retinoblastoma (pRBR) no parece esencial para la replicación del virus (Settlage et al. 2005).

Proteína AC2/TrAP

La proteína TrAP al igual que la mayoría de las proteínas de los geminivirus, es una proteína multifuncional que participa en la transactivación de genes tardíos (*CP* y *BV1*), en la supresión del silenciamiento, y en la supresión de los mecanismos de defensa de la planta (Hao et al. 2003; Sunter & Bisaro 1992; Voinnet et al. 1999; Trinks et al. 2005). La proteína TrAP tienen tres dominios conservados importantes para efectuar sus funciones: un dominio básico con una señal de localización nuclear (SLN), un dominio de unión al ADN con motivo de dedo de zinc, y un dominio activador ácido en el C-terminal de la proteína (Hartitz et al. 1999). El dominio básico con la SLN funcional y el dedo de zinc se requieren para la eficiente supresión del silenciamiento (Wezel et al. 2003; Trinks et al. 2005). Se ha propuesto que la proteína TrAP se une e inhibe a la adenosín-kinasa (ADK). La inhibición de la ADK evita que el ADN viral sea metilado, lo que suprime el silenciamiento (Wang et al. 2003; Bisaro 2006). La proteína también se une e inhibe a la kinasa SNF1, lo que altera el metabolismo de la planta huésped y provoca que sea susceptible al ataque viral (Hao et al. 2003). La función de transactivación de los genes tardíos requiere que TrAP se una a la proteína PEAPOD2. Ésta proteína tiene la capacidad de unirse a los promotores de *CP* (Lacatus & Sunter 2009). Se ha propuesto que la proteína TrAP tiene como blanco un motivo en el ADN

denominado “elemento tardío conservado” (CLE) (Argüello-Astorga et al. 1994), el cual es importante en la transactivación del gen *CP* en el virus *Pepper huasteco yellow vein virus* (Ruiz-Medrano et al. 1999).

Proteína AC4/C4

La proteína AC4/C4 está involucrada en el movimiento de los virus monopartitas y se ha descrito como un determinante de patogénesis en algunos virus monopartitas y bipartitas del VM (Vanitharani et al. 2004; Rojas et al. 2001), ya que plantas transgénicas que expresan la proteína C4 de *Tomato leaf curl virus* presentan síntomas muy parecidos a los producidos por la infección del virus (Krake et al. 1998; Rigden et al. 1994). En adición a las funciones anteriores, se ha demostrado que la proteína AC4 de los virus bipartitas ACMV y SLCMV puede contrarrestar algunos mecanismos de defensa antiviral de las plantas (Vanitharani et al. 2004; Wezel et al. 2002). A diferencia de la proteína AC4/C4 de los virus del VM, a la proteína AC4 de los virus del NM no se le ha encontrado una función, y mutaciones en esta proteína no parecen afectar el desarrollo de la infección o alterar los síntomas en la plantas inoculadas con el virus mutante (Hoogstraten et al. 1996; Pooma & Petty 1996; Fontenelle et al. 2007).

Proteína AC1/Rep

La proteína de replicación es una proteína multifuncional que es esencial para la replicación del virus. Esta proteína reconoce y une secuencias repetidas llamadas iterones, localizadas en la región del origen de replicación (Fontes et al. 1994; Argüello-

Astorga et al. 1994). La proteína cataliza el corte de una de las cadenas del ADN en una secuencia noamérica conservada, la cual se localiza en el “asa” de un elemento con el potencial de formar una estructura de “horquilla”. Este corte endonucleolítico por parte de Rep inicia la replicación del genoma viral por un mecanismo de círculo rodante (Heyraud-Nitschke et al. 1995). La proteína también tiene actividad de helicasa en la fase de elongación (Gorbalenya & Koonin 1993; Pant et al. 2001), y esta función se asocia a la presencia de un motivo llamado P-loop, característico de las proteínas que hidrolizan el ATP (dominio NTP) (Desbiez et al. 1995). La proteína Rep tiene la capacidad de formar homo-oligómeros y de unirse a la proteína REn. Se ha propuesto que la homo-oligomerización es un prerequisite para que la proteína pueda unirse al ADN pero no lo es para producir el corte (Orozco & Hanley-Bowdoin 1998). Como se mencionó anteriormente, las mutaciones que impiden la unión entre REn y Rep disminuyen los niveles de ADN viral (Settlage et al. 2005). La proteína Rep se une a proteínas de la célula huésped implicadas en la replicación del ADN (v.gr: PCNA) (Castillo et al. 2003), y en la división y diferenciación celular alterando el control transcripcional e induciendo la maquinaria de replicación del huésped (Hanley-Bowdoin et al. 2004; Kong et al. 2000).

Proteína BV1/NSP

La principal función de la proteína BV1 es la de transportar el ADN viral hacia dentro y fuera del núcleo. BV1 también se ha visto involucrada en la interferencia de defensas antivirales de la planta huésped, y posiblemente esté involucrada también en la diferenciación celular. La proteína BV1 tiene dos SLN y una señal de exportación

nuclear (NES). Puede interactuar con la importina α y es capaz de unirse a ADN de cadena sencilla y doble de manera no-específica de secuencia, pero sí dependiente del tamaño y forma específica de la molécula de ADN; estas cualidades le permiten exportar e importar el ADN viral del núcleo (Rojas et al. 1998; Kass et al. 2006; Sanderfoot et al. 1996; Ward & Lazarowitz 1999). La actividad de supresión de la defensa antiviral mediada por la proteína BV1 está determinada por la unión a las proteínas-quinasas parecidas a receptores (RLKs), a las que inhibe. Mutantes de RLKs en *Arabidopsis thaliana* son más susceptibles a la infección del begomovirus *Cabbage leaf curl virus* indicando que estas proteínas están involucradas en la defensa antiviral (Fontes et al. 2004). Se ha descrito que la proteína BV1 también puede unirse e inhibir a la proteína acetiltransferasa de *Arabidopsis* (AtNSI), la cual se ha propuesto que acetila proteínas de la planta que inducen la diferenciación celular. Así, la interacción BV1-AtNSI quizás mantenga a las células del huésped en un estado indiferenciado, en el que se encuentran más susceptibles a la infección por los BGVs (McGarry et al. 2003).

Proteína BC1/MP

La proteína BC1 está involucrada en el movimiento sistémico del virus, ya que facilita el movimiento célula a célula, al igual que lo hacen las proteínas que pertenecen a la familia de las proteínas del movimiento (MP). Las proteínas de esta familia de proteínas del movimiento actúan aumentando el tamaño de exclusión de los plasmodesmos, y en algunos casos como el del virus *Cauliflower mosaic virus*, mediante la formación de túbulos en el canal del plasmodesmo (Kasteel et al. 1996; Vaquero et al. 1997). Si bien

las proteínas BC1 no están clasificadas dentro de la familia de las MP, las proteínas BC1 poseen las mismas habilidades que las proteínas de esta familia: son capaces de unirse al ADN de manera tamaño- y forma-específica, pero no reconocen secuencias nucleotídicas particulares; pueden interactuar con los plasmodesmos, y algunas son capaces de asociarse con túbulos derivados del retículo endoplásmico (Ward et al. 1997; Rojas et al. 1998; Lazarowitz & Beachy 1999).

Se han propuesto dos modelos que explican el movimiento célula a célula mediado por BC1, debido a que existen diferencias aparentes entre begomovirus en la capacidad de BC1 de interactuar con el ADN. El primer modelo propone que el movimiento célula-célula se lleva a cabo por la cooperación de las proteínas BV1 y BC1. BV1 forma primero un complejo con el ADN viral, y este último se une luego a BC1. El complejo tripartita BC1:BV1:ADN es transportado después al plasmodesmo, donde BC1 aumenta el tamaño de exclusión del mismo (Fig. 3). Este modelo fue propuesto porque algunas proteínas BC1 de begomovirus, como la de *SLCV*, no tienen la capacidad de unirse a ADN pero sí a la proteína BV1 (Sanderfoot et al. 1996; Sanderfoot & Lazarowitz 1995; Sanderfoot & Lazarowitz 1996). El segundo modelo propone que BV1 saca el ADN del núcleo y lo transfiere a la proteína BC1, la cual lleva entonces el ADN hacia los plasmodesmos. Este modelo se propuso debido a la capacidad que tienen las proteínas de movimiento de los virus BDMV y AbMV, tanto BC1 como BV1, de unirse al ADN (Rojas et al. 1998; Gilbertson et al. 2003). El transporte del complejo nucleoproteico viral del citoplasma a los plasmodesmos posiblemente se lleve a cabo de dos formas diferentes. La primera posibilidad es que el virus se mueva a través de túbulos, puesto que se ha descrito que la proteína BC1 se asocia con túbulos que se

extienden del retículo endoplásmico a los plasmodesmos (Ward et al. 1997). Otra forma potencial de movimiento sería a través de vesículas que se generan en el retículo endoplásmico y que tienen como destino la periferia de la célula. Las proteínas BC1 tienen la capacidad de interactuar con la proteína sinaptotagmina, ésta proteína regula la endocitosis y el reciclamiento de endosomas así como el transporte de proteínas de la planta a los plasmodesmos (Lewis & Lazarowitz 2010).

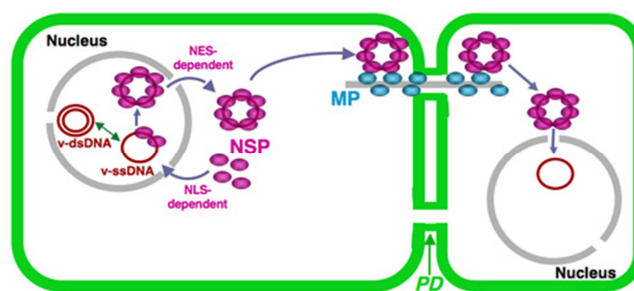


Figura 3. Modelo del movimiento de begomovirus.

El modelo propone que las proteínas NSP y MP interactúan para lograr el movimiento sistémico del virus. V-dsDNA: ADN de cadena doble; V-ssDNA: ADN de cadena sencilla; Nucleus: Núcleo; NES-dependent: dependiente de SEN; NLS-dependent: dependiente de SLN. (Figura reproducida de (Lazarowitz & Beachy 1999).

Tabla 1 Lista de las proteínas de begomovirus

GEN	PROTEÍNA	TAMAÑO	FUNCIÓN
AC1	AC1/Rep	345-361aa	Inicia la Replicación viral.
AC2	AC2/TrAP	130aa	Activa la expresión de los genes tardíos CP, BV1. Supresor del silenciamiento.
AC3	AC3/REn	132aa	Aumenta la replicación viral.
AC4	AC4	87aa	Determinante de patogénesis.
AV1	AV1/CP	254aa	Forma la cápside viral. Proteína lanzadera.
AV2	AV2 (pre-coat)	115aa	Movimiento viral eficiente. Supresor del silenciamiento.
BV1	BV1/NSP	258aa	Movimiento dentro de la célula.
BC1	BC1/MP	293aa	Movimiento entre células.

Satélites

Los satélites son definidos como entidades extracromosómicas que dependen de un virus patrocinator (“helper”) para su replicación y encapsidación. Los satélites no son necesarios para la replicación del virus patrocinator y no presentan una secuencia con identidad significativa al genoma de éste último (Murant & Mayo 1982). La asociación con satélites es una característica común de los virus de ARN y los begomovirus (Dry et al. 1997). Los satélites de BGVs consisten de ADNccs de aproximadamente 1300 pares de bases (Bridson & Stanley 2006). Estos satélites están asociados principalmente a los BGVs monopartitas (Dry et al. 1997; Nawaz-ul-Rehman & Fauquet 2009), aunque recientemente se ha descrito la asociación de satélites con BGVs bipartitas del Nuevo Mundo (Romay et al. 2010; Paprotka et al. 2010). Los BGVs se asocian con tres tipos de satélites: los alfa-satélites (ADN α), los beta-satélites (ADN β) y el satélite asociado al virus *Tomato leaf curl virus* (ToLCV-Sat ADN).

Alfa-satélites

Los ADN α son satélites de un tamaño equivalente a la mitad del de un componente genómico de los begomovirus, y presentan una organización genómica característica, que incluye a una región de 200 pb rica en adeninas y un origen de replicación que incluye un nona-nucleótido diferente al de los begomovirus (TAGTATT/AC). Estos satélites codifican una proteína Rep similar a la de nanovirus, la cual les permite replicarse de modo independiente del virus patrocinator. Sin embargo, los ADN α requieren del componente A de los BGVs para su movimiento, encapsidación y transmisión mediada por el vector (Saunders & Stanley 1999). Los ADN α

frecuentemente se asocian con los complejos formados por los satélites β y los BGVs (Nawaz-ul-Rehman & Fauquet 2009). Estas tres entidades no comparten similitud en la organización genómica excepto por la estructura tallo-asa necesaria para la replicación (Patil & Fauquet 2010).

Beta-satélites

Los llamados beta-satélites o ADN β contienen: a) un marco de lectura abierto (ORF) en sentido (-) que codifica para la proteína β C1, b) una región rica en adeninas, c) una región conservada entre los satélites, de ~ 200 pb y, d) una estructura tallo-asa con una secuencia nonamérica idéntica a la de los geminivirus (TAATATT/AC), también involucrada en la replicación (Fig. 1E) (Bridson & Stanley 2006). Los ADN β aumentan la eficiencia de infección del virus patrocinator en la planta, ayudan a la acumulación del ADN viral y se sabe que están involucrados en el desarrollo de síntomas (Patil & Fauquet 2010; Saeed et al. 2005; Saunders et al. 2000), provocados tal vez por la actividad supresora del silenciamiento de la proteína β C1 (Yang et al. 2011; Cui et al. 2005; Saeed et al. 2005). En condiciones experimentales los ADN β tienen la capacidad de complementar las funciones de movimiento del ADN-B de los virus bipartitas (Patil & Fauquet 2010; Saeed et al. 2007).

Otros satélites: ToLCV-Sat

El último tipo de satélite, el ADN circular asociado a ToLCV, sólo se ha encontrado hasta ahora en infecciones causadas por virus *ToLCV* en Australia. El ADN de este

satélite es de solo 682 pb (un cuarto del genoma begomoviral típico). Como todos los satélites, su ADN no presenta identidad en secuencia significativa con los begomovirus, excepto por el nanonucleótido TAATATTAC encontrado en la estructura tallo-asa. Este satélite depende enteramente del virus patrocinator para replicarse, moverse y encapsidarse, no contiene ningún ORF discernible, y no se ha involucrado en el ciclo infeccioso del begomovirus (Dry et al. 1997). Se ha sugerido que este satélite representa un ADN- β defectuoso (Saunders et al. 2000).

Origen de los geminivirus

El origen y la antigüedad de los geminivirus son cuestiones que no han sido resueltas de un modo definitivo todavía, aunque se han propuesto varias hipótesis que intentan resolverlas. Una hipótesis propone que los geminivirus son descendientes de fagos de ADNccs como PhiX174, sobre la base de la similitud que existe entre las proteínas de replicación de estos sistemas virales, y porque ambos se replican por un mecanismo de CR (Rybicki 1994). Una observación adicional que apoya la hipótesis mencionada es la de que el promotor *CP* de algunos geminivirus es funcional en *E. coli* (Petty et al. 1986).

Otra serie de hipótesis postula que los geminivirus son descendientes de plásmidos (Rojas et al. 2005). Entre los plásmidos que podrían haber dado origen a los geminivirus destacan algunos de origen bacteriano, otros de algas rojas y algunos de fitoplasmas. La noción de un origen en plásmidos bacterianos surgió del hallazgo de una similitud significativa entre las proteínas de replicación de geminivirus y las de una familia de plásmidos de bacterias Gram positivas (Koonin & Iiyina 1992). Observaciones

adicionales que indican que los geminivirus poseen características de plásmidos de procariontes, son sus ARN mensajeros policistrónicos y la capacidad que tienen de replicarse en células procariotas (Rigden et al. 1996; Selth et al. 2002; Frischmuth et al. 1990). La única evidencia que sustenta la hipótesis de que los geminivirus se originaron a partir de plásmidos de algas rojas es la similitud entre una región de 55aa de la proteína Rep del mastrevirus *Wheath dwarf virus* y la proteína de tipo Rep de un plásmido del alga roja *Pophyra pulchera*, las cuales comparten una similitud mayor al 31% (Nawaz-ul-Rehman & Fauquet 2009). Por otra parte, la hipótesis que postula que los geminivirus se originaron de plásmidos de fitoplasmas se basa en el resultado de la búsqueda de homólogos de la proteína Rep de geminivirus en la base de datos de NCBI, mediante la herramienta PSI-BLAST. La búsqueda mostró que los mejores “scores” se obtienen con las proteínas Rep de algunos plásmidos de fitoplasmas (Krupovic et al. 2009). Sin embargo, poco después Saccardo et al. (2011) refutaron esa hipótesis al mostrar que la proteína Rep de los plásmidos de fitoplasmas fue probablemente adquirida por recombinación con un geminivirus. Todas las hipótesis anteriores difieren en el punto relativo al origen de la proteína de replicación, pero coinciden en la noción de que la CP fue adquirida a través de un evento de recombinación. Una conjetura plausible es que la proteína CP fue reclutada de un virus de plantas, como el *Satellite tobacco necrosis virus* (Krupovic et al. 2009) o de un virus relacionado con sistemas virales presentes en ambientes marinos (Saccardo et al. 2011).

En relación a la escala temporal las estimaciones respecto al origen de los geminivirus son extremadamente variables. Por ejemplo, Nawaz-ul-Rehman & Fauquet

(2009) estiman que si los geminivirus se originaron realmente de plásmidos de algas rojas, es posible que el ancestro de los geminivirus haya aparecido hace 450 millones de años. Esta estimación resulta incongruente con la historia evolutiva de las plantas superiores, que aparecieron hace menos de 150 Ma. La mayoría de los especialistas coinciden en afirmar que los geminivirus probablemente evolucionaron en el Cretácico, cuando las plantas angiospermas aparecieron y se diversificaron, y hubo además una explosión evolutiva de los insectos vectores de los geminivirus, hace unos 130 Ma (Willis & McElwain 2002).

Planteamiento de los problemas que se abordan en el estudio.

Los datos disponibles han permitido estimar, de manera un tanto amplia e imprecisa, el periodo en el cual evolucionaron los geminivirus. Sin embargo, varias cuestiones importantes relativas a la historia evolutiva de los begomovirus no han sido resueltas de manera concluyente todavía, como las siguientes:

- 1) ¿En qué región geográfica evolucionaron los primeros BGVs? En otras palabras, ¿cuál fue el centro de origen de este subgrupo de los geminivirus?

- 2) ¿De dónde provienen los BGVs del Nuevo Mundo, que se distinguen tan claramente de los BGVs de otros continentes? ¿Hubo acaso más de un evento de introducción de estos virus al continente americano en tiempos prehistóricos?

- 3) ¿Cómo se originó el segundo componente genómico de los begomovirus, el cual no tiene homólogos en ningún otro género de los *Geminiviridae*?

Objetivo del Estudio

El objetivo global de este trabajo es reconstruir algunos aspectos fundamentales de la historia evolutiva de los BGVs, partiendo de la idea de que es factible hacerlo si se realiza un análisis integral de los datos genómicos, filogenéticos y paleobiogeográficos disponibles.

Objetivos Específicos

1. Inferir las características genómicas del ancestro común de los begomovirus a partir del análisis comparado de las secuencias codificantes y reguladoras de las especies conocidas.
2. Definir el posible Centro de Origen de los BGVs a partir de los datos paleobiogeográficos de su vector y los de las plantas con fósiles moleculares de BGVs en sus genomas, así como de la actual distribución geográfica de los linajes principales y la identificación de “firmas” biogeográficas o específicas de estirpe en las proteínas virales, todo ello analizado en el contexto de la fragmentación de supercontinentes en el período inicial de la historia evolutiva de los BGVs.
3. Resolver la cuestión del origen de los BGVs del Nuevo Mundo, que representa en apariencia el linaje más reciente y diferenciado de este género de la familia *Geminiviridae*.
4. Abordar la cuestión del origen evolutivo del segundo componente genómico de los begomovirus, utilizando una amplia gama de metodologías comparativas para establecer homologías remotas.

RESULTADOS-PARTE I

Centro de Origen de los begomovirus y organización genómica del ancestro común.

I.1- Filogenia de los componentes genómicos de begomovirus

Los estudios de genómica comparada nos permiten comprender cómo cambian los genomas de las entidades biológicas en el tiempo. El procedimiento básico consiste en comparar los genomas de diferentes especies para descubrir similitudes y diferencias entre ellos. Las comparaciones primero se realizan entre un grupo de especies relacionadas filogenéticamente y luego el grupo estudiado puede compararse con un grupo de especies más distantes por lo tanto, es necesario construir un árbol filogenético que nos ayude a identificar los grupos de especies relacionadas que pueden compararse de modo más productivo para los fines específicos de la investigación.

Para deducir las relaciones evolutivas de los componentes genómicos A y B de BGVs, se realizó un alineamiento múltiple, con el algoritmo ClustalW, de las secuencias del ADN-A de 63 BGVs, por un lado, y las secuencias completas del ADN-B de 58 especies virales, incluyendo al mastrevirus *MSV* como grupo externo. Los árboles filogenéticos se reconstruyeron usando el método Bayesiano incluido en el software Mr. Bayes 3.2 (Ronquist et al. 2012) (Fig. 4).

La reconstrucción filogenética basada en la secuencia completa del componente A mostró que los BGVs se dividen en dos grandes grupos, uno que comprende a los

BGVs del llamado Viejo Mundo (Africa, Europa, Asia, y Oceanía) y el otro lo comprenden los BGVs del Nuevo Mundo (NM) (Fig. 4A). El grupo de los virus del NM lo integran los BGVs nativos del continente americano, todos ellos con genomas bipartitas. En este grupo existe un linaje especial llamado “clado del SLCV”. Los miembros de este clado presentan diferencias bien definidas con el resto de los begomovirus en la secuencia de aa de su proteína Rep y en la composición y organización del origen de replicación (Argüello-Astorga et al. 1994). El grupo de BGVs del VM lo conforman virus bipartitas y monopartitas pertenecientes a distintos linajes. Uno de estos linajes es el de los “Legumovirus”, virus que infectan leguminosas y son nativos de la India, el Sureste Asiático, y África. La segunda estirpe la conforman los virus nativos de Asia y Australasia, y el tercer subgrupo se encuentra integrado por los BGVs originarios de África, el Oriente Medio y la región mediterránea de Europa.

La filogenia derivada de la secuencia completa del componente B también divide a los BGVs en virus del NM y virus del VM (Fig. 4B). Las relaciones filogenéticas derivadas de la comparación de los componentes B de los virus del NM son, en la mayoría de los casos, congruentes con las obtenidas en el análisis comparado de los componentes A; sin embargo, los cladogramas de los ADN-A y -B de los virus del VM muestran algunas diferencias importantes en las relaciones filogenéticas inferidas. Por ejemplo, en la filogenia del ADN-A el linaje de los Legumovirus (v.gr: MYMV y SbCBV) parece estar más estrechamente relacionado a los BGVs de América que a los BGVs de África, en tanto que en el cladograma del ADN-B los Legumovirus se agrupan claramente con los BGVs africanos (Fig.4). Más importante todavía es la diferencia observada en el cladograma del componente B, en el que los BGVs del VM se dividen

en 2 clados principales, mientras que en el árbol filogenético de los ADN-A de esos BGVs se distinguen 3 linajes mayores: el de los Legumovirus, el asiático y el africano. Las filogenias derivadas de las secuencias de aminoácidos de las proteínas BV1 y BC1, codificadas en el ADN-B, presentan la misma topología que el árbol filogenético obtenido con los componentes genómicos B completos (Fig. 5), lo que sugiere que ambas proteínas han evolucionado coordinadamente desde la emergencia evolutiva del segundo componente genómico.

Es importante destacar que el árbol filogenético construido a partir de la secuencia completa del ADN-B de los BGVs, comparado con la filogenia derivada a partir del ADN-A, presenta un mayor soporte en los nodos de las ramas. Al considerar esa mayor robustez de la filogenia derivada del ADN-B, decidimos utilizarla como la principal guía para reconstruir la historia evolutiva de los BGVs. Esta filogenia muestra la existencia de tres “Linajes Primarios” de BGVs, los cuales están soportados por sus probabilidades posteriores mayores al 94% (Fig. 4B).

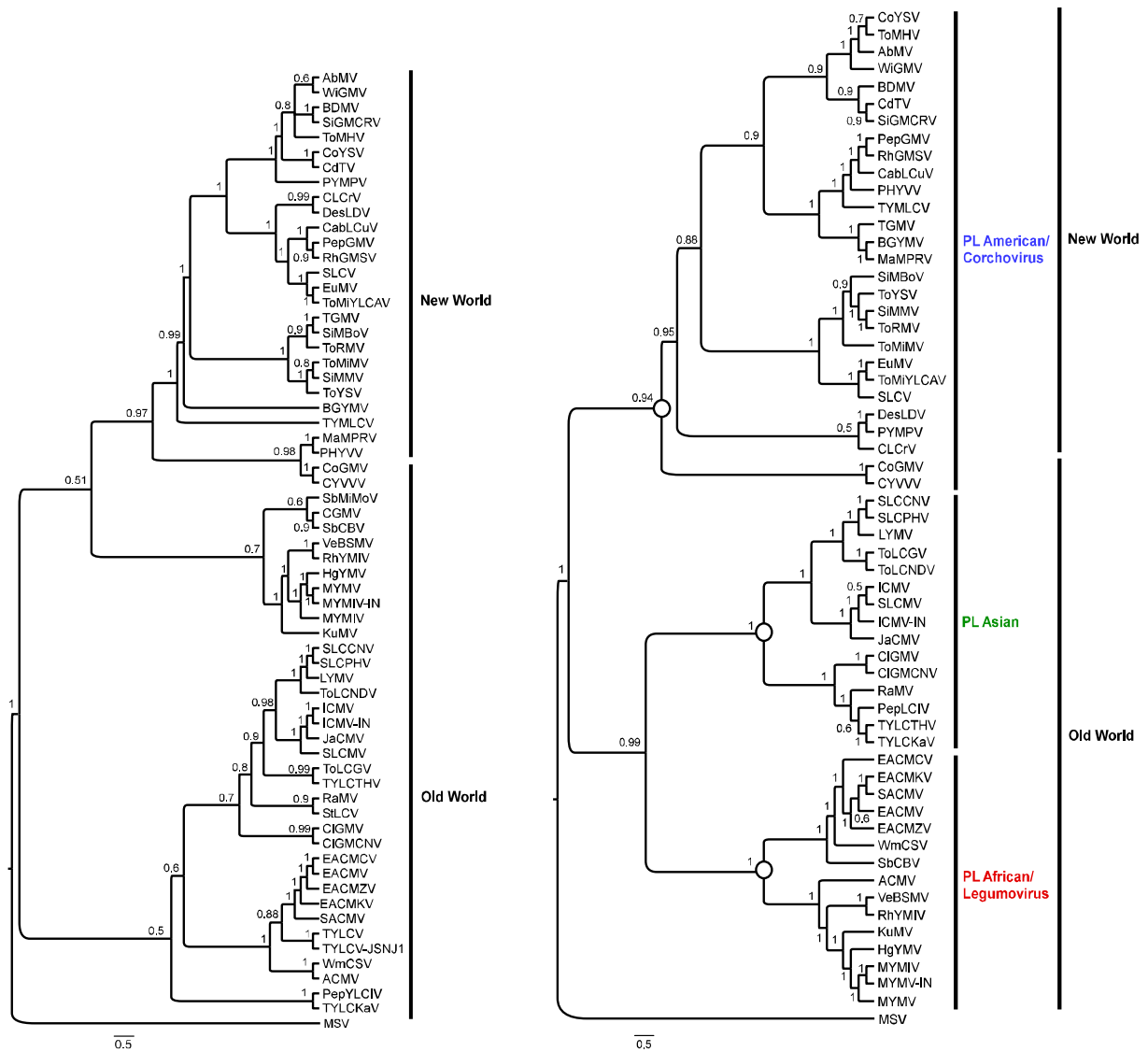


Figura 4. Reconstrucción filogenética de los componentes A y B de begomovirus.

Los árboles filogenéticos están basados en el alineamiento de 63 secuencias nucleotídicas de componentes genómicos A (Panel A) y 58 secuencias del componente genómico B (Panel B). La reconstrucción filogenética se realizó con el método bayesiano usando el programa Mr. Bayes 3.2, con el modelo GTR+G+I. La escala hace referencia a los cambios esperados por sitio. Los árboles fueron enraizados usando al mastrevirus MSV como grupo externo. El número asociado a cada nodo de los árboles indica el soporte de cada clado y se encuentran expresados en probabilidades posteriores. En el panel B, El árbol del ADN-B se divide en tres linajes primarios (LPs): Americano/Corchovirus, Asiático and Africano/Legumovirus. Los círculos abiertos representan los nodos de cada uno de los LPs presentes en el árbol filogenético de ADN-B. El código de color de los LPs los usaremos en siguientes figuras con el fin de distinguir a que linaje pertenece cada virus usado. Los números de acceso y los acrónimos usados para cada una de las secuencias están enlistados en la Tabla 2.

Los tres linajes primarios (LPs) de BGVs fueron denominados de acuerdo a las estirpes virales que incluyen. 1) LP América/Corchovirus, que comprende a todos los BGVs nativos del NM y dos especies nativas de Vietnam y la India, *Corchorus yellow vein vietnam virus* (CYVVV) y *Corchorus golden mosaic virus* (CoGMV), las cuales infectan plantas de yute (*Corchorus capsularis*) (Ha et al. 2006; Ha et al. 2008)08), razón por la cual han sido denominados “Corchovirus”. El análisis de sus secuencias muestra que estos virus son filogenéticamente más cercanos a los BGVs de América que a otros virus asiáticos (Fig. 4). 2) LP África/Legumovirus, que incluye a los BGVs del continente africano y al linaje de los “Legumovirus”, que infectan plantas leguminosas y se distribuyen tanto en Asia (India, Thailandia, Vietnam) como en África (Nigeria) (Fig. 4). El tercer LP, denominado simplemente “Asiático”, está integrado por todos los BGVs de Asia y Oceanía que no pertenecen a los linajes de Corchovirus y Legumovirus. En este LP se distinguen dos grandes ramas: los virus aislados en el subcontinente Indio (ICMV y parientes), y los virus de China, el SE asiático y Australasia (Fig.4).

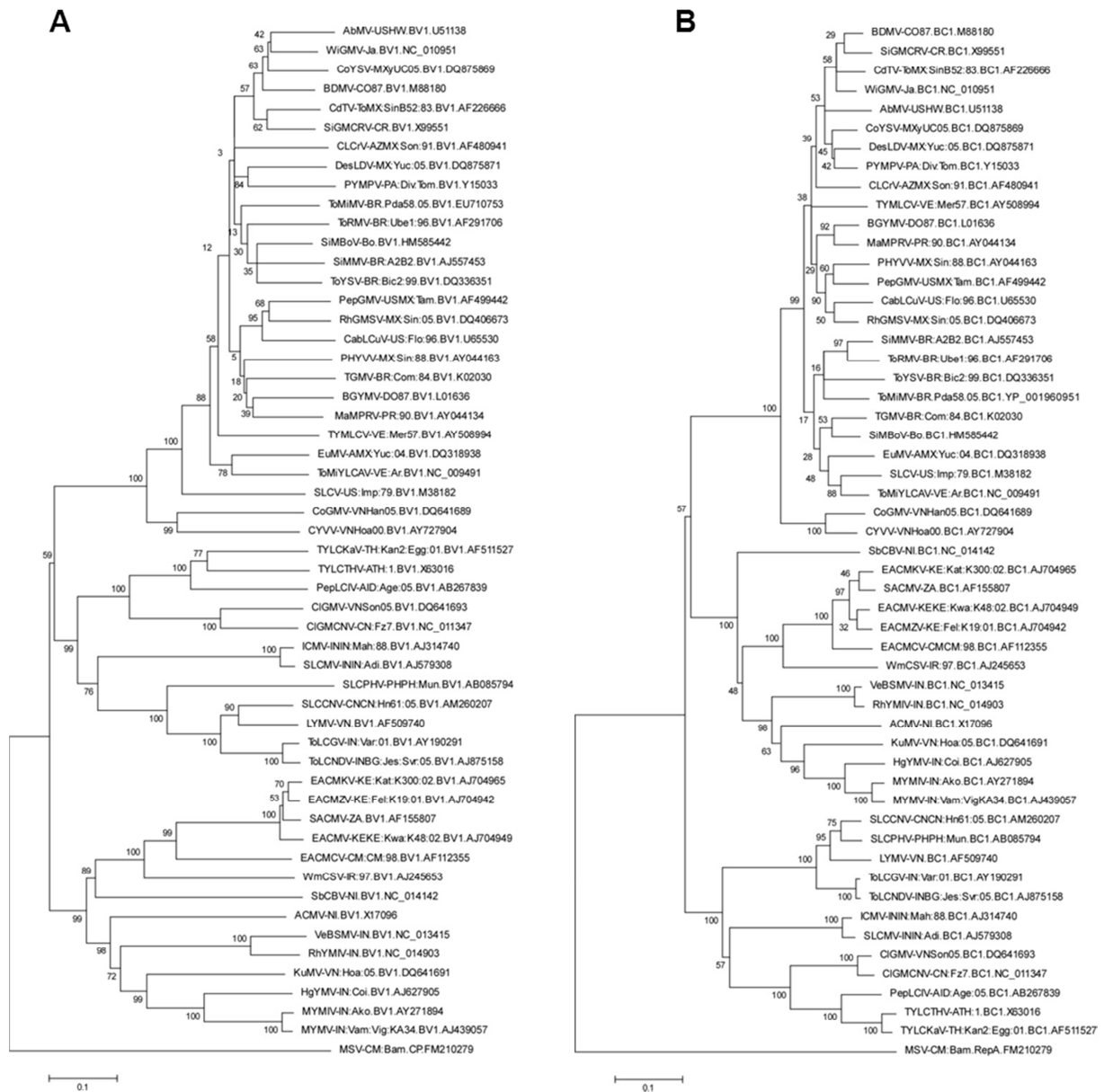


Figura 5 Reconstrucción filogenética de las proteínas BV1 y BC1 de begomovirus.

Los árboles filogenéticos están basados en el alineamiento de 53 secuencias de aminoácidos de proteínas BV1 (Panel A) y 53 secuencias de la proteínas BC1 (Panel B). La reconstrucción filogenética fue inferida usando el método de Neighbor-Joining usando el programa MEGA 5. La distancia evolutiva se obtuvo mediante el método de número de diferencias y las unidades representan el número de bases diferentes por secuencias. Los árboles fueron enraizados usando como grupo externo la proteína CP del mastrevirus MSV en el caso del árbol de BV1 y usando la proteína RepA del mismo virus en el caso de BC1. El número asociado a cada nodo de los árboles indica el soporte de cada clado y representan el porcentaje obtenido del análisis bootstrap (1000 replicas). Los acrónimos usados para cada una de las secuencias están enlistados en la Tabla 2.

I.2 - Identificación de marcadores biogeográficos y “firmas” específicas de linaje en las proteínas BV1 y BC1.

La identificación de secuencias cortas de aminoácidos (aa) altamente conservadas en un grupo de proteínas relacionadas es una eficiente forma de recuperar y clasificar los genes que pertenecen a un mismo linaje genético (Vieira et al. 2008). Consecuentemente, decidimos analizar las proteínas codificadas en el componente B en busca de motivos característicos de estirpes biogeográficas o específicas de los LPs definidos previamente. También realizamos el análisis de la proteína de la cápside, a fin de establecer si los motivos biogeográficos o específicos de LP presentes en una proteína codificada por el DNA-A correlacionan con los observados en las proteínas BV1 y BC1.

I.2.1 - Análisis de las proteínas BV1

Las proteínas BV1 de las especies de BGVs analizados en la etapa previa fueron ordenadas en grupos correspondientes a los tres LPs, o alternativamente, en grupos de virus nativos de los diferentes continentes. Los consensos obtenidos del alineamiento de estos grupos de proteínas fueron luego comparados de modo sistemático para identificar motivos (“motifs”) distintivos de linajes biogeográficos o de LP. Este análisis condujo a la demarcación de varios motivos que satisfacen en mayor o menor grado el criterio de “firma” (“signature”) característica de un Linaje Primario. Un dominio de BV1 que contiene ejemplos destacados de esas “firmas” específicas de LP es el segmento 133-147 de la proteína de los BGVs del continente americano, y su equivalente en los BGVs asiáticos (posiciones ~145-160) y los miembros del LP Africa/Legumovirus (~130-

145). Como se puede apreciar en la figura 6 los motivos o “firmas” encontradas en BV1, son exclusivas de cada uno de los tres LP. Por ejemplo la firma del LP America/Corchovirus [V***DRKPHV***GXSGSLHT] está presente en la proteína BV1 de todos los virus nativos del NM (v.gr: CdTV y SLCV) y en los Corchovirus de Vietnam y la India (v.gr: CYVV), lo que apoya la noción de un origen común de estas dos estirpes geográficamente disyuntas. Los virus del LP Asiático (v.gr: TYLCKaV y PepYLCIV del SE asiático, y ToLCNDV de la India) tiene una “firma” con un consenso más degenerado [R**DKK***PCXF*S(A/S)VDPLIP] pero claramente distintivo del LP, y algo análogo se observa en el caso de la “firma” equivalente [V***DK*****K***PF*V*PE*GvNXLP(T/S)] del LP Africa/Legumovirus que incluye BGVs de dos continentes. Las firmas identificadas en el segmento ~130-145 de BV1 (en las que V*=V, L, I ó M; K*= K ó R; F*= F ó Y; E*= E ó D) apoyan fuertemente la idea de que la clasificación de los BGVs bipartitas en 3 LPs refleja apropiadamente la división natural de esos virus.

Linaje “Americano/Corchovirus”	
133-VDRK PHV SQ-SGRLHT-147	SLCV-México
133-VDRK PHLGA -SGSLHT-147	CdTV-México
130-MDRK PHVGP -SGSLPK-144	CYVV-Vietnam
Linaje “Asiático”	
155-RDKR PCQFSAVDPIIP -170	TYLCKaV-Tailandia
147-RDKK PCFESSVDPLIP -162	PepYLCIV-Indonesia
145-RDK SPKIYSATQPLIP -160	ToLCNDV-India
Linaje “Africano/Legumovirus”	
132-LDDK PYLPEGVNKLPS -147	MYMIV-India
130-LDRK PYVDPGANELPS -145	VeBSMV-India
130-VDKR PVPEGVNTLPT -145	EACMCV-África
134-VDKK PVPEGV SILPT-149	SACMV-África

Figura 6. Motivo 130-145 de la proteína BV1.

Los virus utilizados en esta figura son representativos de cada uno de los LPs, los colores utilizados para distinguir los LPs son los mismos que los usados en la figura 4. Los residuos conservados tan solo en un LP están resaltados a color. Los números de acceso y los acrónimos de las proteínas pueden consultarse en la Tabla 3.

I.2.2-Análisis de las proteínas BC1

El análisis comparativo de las proteínas BC1 hipotéticas de todas las especies begomovirales analizadas en este estudio nos permitió identificar dos motivos conservados, uno comprendido entre los aminoácidos 80 y 94, y el otro entre los residuos 173 y 191 (coordinadas referidas a los BGVs del NM) (Fig. 7). Al igual que el motivo que se encuentra entre los aa 130-145 de la proteína BV1, las firmas son características de cada uno de los 3 LP. En el motivo comprendido entre los residuos 80-94 del LP America/Corchovirus identificamos cinco residuos aa (en color azul) distintivos de los BGVs de América y los Corchovirus. En el LP Asiático también identificamos cinco residuos aa distintivos, cuatro de los cuales son compartidos entre los virus de la India y los del Sudeste y Centro de Asia. En los virus del LP Africano/Legumovirus identificamos seis residuos diferenciales, de los cuales cuatro son compartidos entre las dos estirpes virales geográficamente disyuntas que conforman el LP (Figura 7a). En el motivo que abarca los aminoácidos 173 al 191 cinco aa son compartidos por los dos grupos que integran al LP America/Corchovirus, dos aa compartidos entre los grupos del LP Asiático y cuatro residuos aa compartidos entre los grupos que conforman el LP Africano/Legumovirus (Figura 7b). La identificación de estos dos motivos en la proteína BC1, en conjunto con el motivo de la proteína BV1 antes descrito, apoya la clasificación de los BGVs en los tres LPs definidos por la filogenia construida a partir del ADN-B.

A		B	
<u>Linaje "Americano/Corchovirus"</u>		<u>Linaje "Americano/Corchovirus"</u>	
80-RMTDNESLQATWTFP-94	DeLDV-México	173-DFSHVDYGRWERKPIRCAS-191	
80-RMTDNESLQASWTFP-94	BGYMV-México	173-DFSHVDYGRWERKLIIRSAS-191	
80-RMNDNESLQTTYTFP-94	CoGMV-Vietnam	173-DFSHVEYGKVERKLIKAS-191	
<u>Linaje "Asiático"</u>		<u>Linaje "Asiático"</u>	
77-RLHEGDSKQAEFTIP-91	TYLCKaV-Tailandia	169-DFWHVPQSKVSRKPVQALS-187	
77-RLHEGDSKQAEFTIP-91	PepYLCIV-Indonesia	169-DFWTVPHSQVSRKPVQGLT-187	
82-RLTEGDTKQAEFTIP-96	ToLCNDV-India	174-DFWSVGSKPQTRRLVDGSR-192	
<u>Linaje "Africano/Legumovirus"</u>		<u>Linaje "Africano/Legumovirus"</u>	
79-RLS YEQAAQAFTFP-93	MYMIV-India	172-DFWSVEKPKPIRRLLNPGP-190	
79-RLCDDEAAQAFTFP-93	VeBSMV-India	172-DFWSVGRPKPVRRLLNPNP-190	
78-RVSL EDAQAFTFP-92	EACMCV-África	171-DFWSVEKGETRRRLLNPTP-189	
78-RVSLDDAAQAFTFP-92	SACMV-África	171-DFWSVEKGETRRRLLNPTP-189	

Figura 7. Motivos específicos de linaje de la proteína BC1.

El motivo 80-94 de BC1 (A) y el motivo 173-191 (B) son de utilidad para identificar LP. Los virus utilizados en esta figura son representativos de cada uno de los LPs, los colores utilizados para distinguir los LPs son los mismos usados en la figura 5. Los aminoácidos a color representan los residuos distintivos de cada LP. Los aminoácidos compartidos entre dos grupos que integran un LP se presentan en color y en negritas, y los residuos que pertenecen solo a un grupo del LP son coloreados y normales. Los números de acceso y los acrónimos de las proteínas pueden consultarse en la Tabla 3.

1.2.3 Análisis de las proteínas CP

El análisis comparativo de la proteína de la cápside codificada por los BGVs de los tres LPs nos permitió identificar dos dominios de la proteína que contienen motivos que se conservan en linajes específicos, algunos de los cuales también constituyen marcadores biogeográficos excepcionalmente robustos. El primer dominio se localiza en el extremo N-terminal de CP, y comprende los residuos aa 1-16, y el segundo dominio se ubica en la región C-terminal de la proteína, entre los residuos 183 y 200 (coordenadas referidas a los BGVs de África) (Fig. 8). En el dominio N-terminal de todos los BGVs del NM se encuentra la firma biogeográfica PWRsMaGT (residuos en color azul, en la figura 8A) previamente descrito por Harrison y colaboradores (2002), que también se conserva en la proteína CP de las dos especies de Corchovirus del SE

asiático, CYVV y CoGMV. Ningún otro BGV examinado presenta este motivo, que constituye por ende una firma exclusiva de los BGVs del LP América/Corchovirus. En el dominio homólogo de la proteína CP de los Legumovirus identificamos una firma específica de linaje, RnF/YDTAFS (recuadro y aminoácidos distintivos en color rojo oscuro de la figura 8A). Los BGVs de Africa y Asia no presentan una secuencia distintiva en la región homóloga de su CP. En el dominio C-terminal de la proteína identificamos una firma exclusiva de las especies virales aisladas en África, Madagascar y las vecinas islas del Índico (aminoácidos rojo brillante de la figura 8B “GGPSGM”). En el resto de los BGVs la secuencia equivalente es GGQYAS (aminoácidos en color azul oscuro de la figura 8B). La combinación específica de los dos motivos distintivos en la proteína CP (Fig. 8C) es consistente con la filogenia derivada del componente A de los BGVs. Por ejemplo, la combinación del motivo N-terminal PWRsMaGT y el motivo GGQYAS/C en el dominio C-terminal distingue a los BGVs del LP América/Corchovirus, en tanto que la combinación de las firmas “NFDTAF” y “GGQYAS/C en las posiciones equivalentes caracteriza a la proteína de los Legumovirus y los distingue de otros BGVs de su mismo Linaje Primario, tal como se observa en el árbol filogenético derivado del ADN-A (Fig. 4A). Estos resultados muestran que la identificación de motivos diferencialmente conservados es una forma eficiente de clasificar proteínas que pertenecen a un mismo linaje genético.

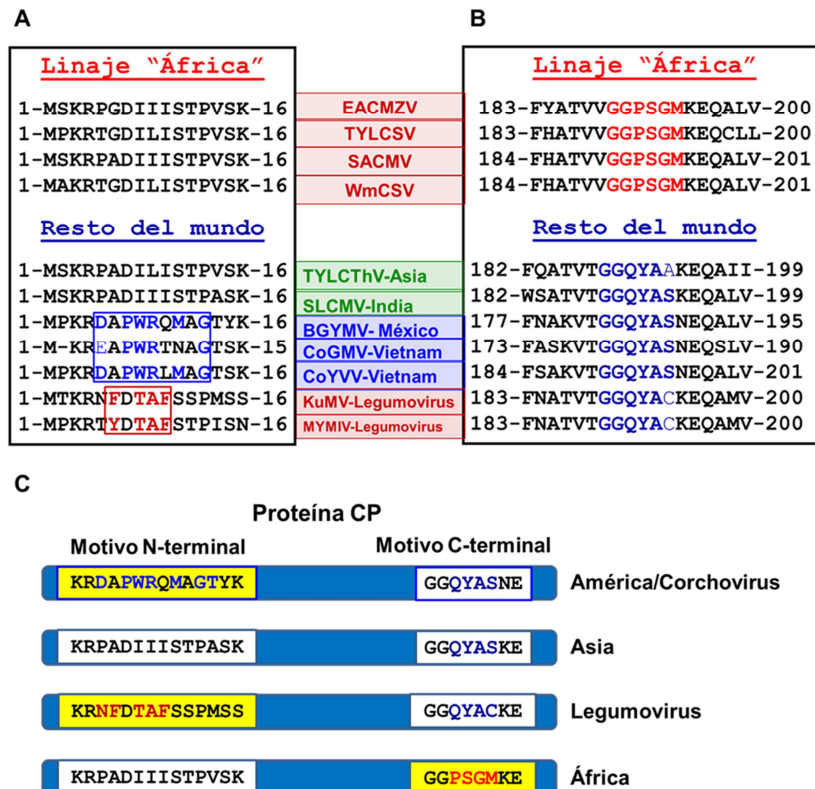


Figura 8. Análisis comparativo de la proteína CP

Se identificaron dos motivos diferenciales en la proteína CP, uno en la región N-terminal (Panel A) y otro en la región C-terminal (Panel B). En el dominio N-terminal se identificó una firma que caracteriza a los virus del LP América/Corchovirus (recuadro y residuos en color azul), y una firma diferente en el dominio homólogo caracteriza a los virus de linaje de los Legumovirus (recuadro y aminoácidos en color rojo oscuro). En el dominio C-terminal se localiza la firma exclusiva de los virus de África (residuos en rojo brillante "GGPSGM") y la que comparten los begomovirus del resto del mundo (residuos en azul oscuro "GGQYAS"). El panel C ilustra las dos clases de motivos mapeados en la proteína CP y muestra que la combinación de motivos N- y C- terminales es única para cada estirpe viral. Los acrónimos de los BGVs utilizados en las comparaciones tienen el color que representa al LP al cual pertenecen. Los nombres y números de acceso son listados en la tabla 3.

I.3- Análisis comparado de las regiones no codificantes

Las secuencias genómicas no codificantes de los BGVs se restringen generalmente a la región intergénica, que contiene el origen de replicación y los promotores de los genes *AC1* y *V2/CP*, en el caso del ADN-A, o de los genes *BC1* y *BV1*, en el del ADN-B. Si bien el análisis de los promotores rara vez se utiliza como indicador de relaciones filogenéticas, por su mayor variabilidad en secuencia respecto a las regiones codificantes, se ha demostrado en numerosos casos que las secuencias reconocidas por factores transcripcionales específicos conservan su posición relativa dentro del promotor, lo que hace factible establecer homologías entre elementos de una serie de promotores de genes que han divergido evolutivamente por períodos extensos. Esta conservación en posición relativa se revela a través de la identificación de “Arreglos Modulares Conservados” (CMAs, por sus siglas en inglés), que se definen como un grupo de “huellas filogenéticas” (“phylogenetic footprinting”) cuya combinación específica, espaciamiento y orientación relativa, se conserva en una serie evolutiva de regiones promotoras (Argüello-Astorga & Herrera-Estrella 1996; 1998). Por ejemplo, el análisis comparado de los CMAs en los promotores de genes *rbcS* (que codifican la subunidad pequeña de Rubisco) ha revelado la conservación de algunos CMAs en plantas que han divergido por más de 400 Ma, como es el caso de musgos, psilofitas, helechos, gimnospermas y plantas con flores (Argüello-Astorga & Herrera-Estrella 1998).

Para los fines de este estudio, decidimos analizar los promotores de los genes *BC1*, *BV1* y *CP* por el método de análisis “filogenético-estructural”, el cual

permite identificar CMAs y otras regiones reguladoras homólogas con diversos grados de divergencia en su secuencia nucleotídica (Argüello-Astorga & Herrera-Estrella 1996). En el caso de los promotores del gen *CP*, cuyo análisis filogenético-estructural fue realizado previamente por Gómez-Castañón (2004), actualizamos el análisis con la inclusión de cerca de 90 nuevas especies de BGVs descritas en los últimos 8 años.

1.3.1. Análisis comparativo del promotor *BV1*

El análisis de la región promotora de BV1 nos permitió identificar CMAs en algunos linajes virales, los cuales mostraron ser muy heterogéneos entre estirpes separadas, por lo que no pudimos reconstruir relaciones evolutivas de manera consistente. Los CMAs encontrados en el promotor de BV1 son específicos de algunos subgrupos dentro de los LPs. Esta heterogeneidad de los CMAs presentes en los promotores BV1 se debe a varias causas. Uno de ellas es el hecho de que los promotores homólogos varían notablemente en longitud, v. gr. los promotores de los virus relacionados TYLCThV y CIGMV presentan una diferencia de 167 nucleótidos en longitud (ver Fig. 4B). Otra razón aparente para la heterogeneidad de los promotores BV1 es el hecho de que algunos de estos presentan inserciones de material genético proveniente del ADN-A (Gregorio-Jorge et al. 2010). Si bien no pudimos establecer relaciones filogenéticas con los CMAs del promotor BV1, es posible que estos elementos compuestos constituyan complejos reguladores de la transcripción del gen BV1 y/o del gen BC1. En el 2005 Shivaprasad y col. delimitaron una secuencia de ~250pb corriente arriba del gen BV1 de MYMV que es capaz de regular tanto la expresión de BV1 como de BC1. Es probable que la secuencia responsable de la regulación de ambos genes del ADN-B

de MYMV sea un CMA adyacente a la horquilla que está conformado por tres sitios de unión de factores de transcripción (Caja G, CLE y Box II Chs) (Fig. 9A). Para investigar la función de este CMA en la regulación génica, se sintetizaron oligonucleótidos complementarios con la secuencia correspondiente a dicho CMA, y el ADN sintético de doble cadena fue insertado arriba de un promotor mínimo (-46/1) 35S en el vector de expresión pBI46S (ver Materiales y Métodos) (Fig. 9B). Esta construcción será evaluada posteriormente en ensayos de expresión transitoria en protoplastos de tabaco NT1, a fin de establecer si funciona como un potenciador de la expresión del gen reportero.

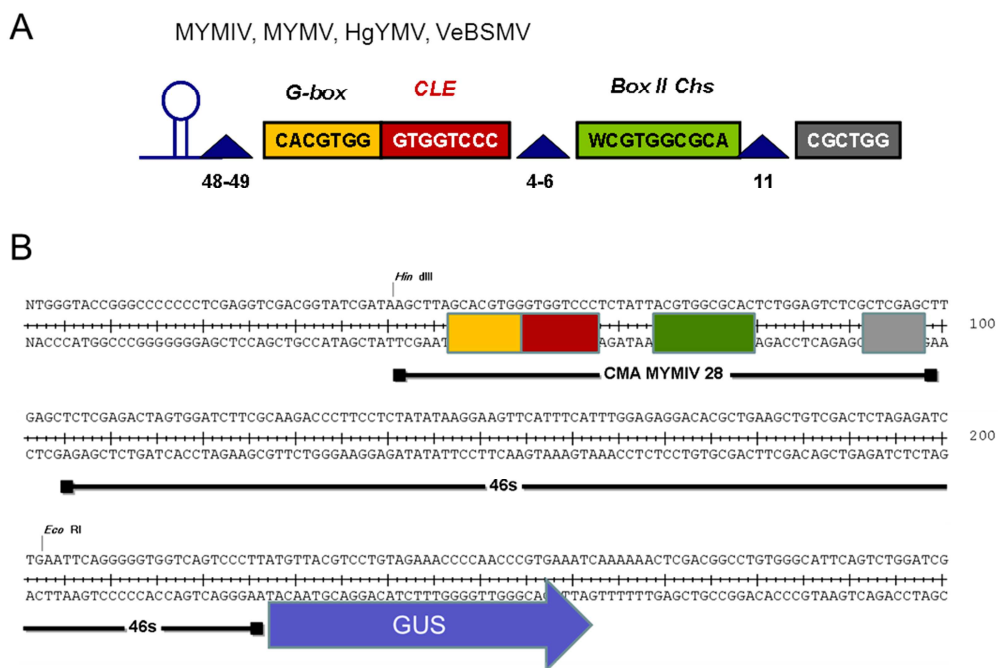


Figura 9. Arreglo modular conservado de Legumovirus

En el panel A se muestra la estructura del CMA adyacente a la horquilla de Legumovirus. El CMA está formado por los sitios de unión de factores de transcripción G-Box, CLE y Box II Chs. El panel B muestra el módulo que se mandó sintetizar y que contiene al CMA de Legumovirus clonado en el vector PBS-46s Gus, el CMA se clonó en el sitio *Hind*III. Los acrónimos pueden ser consultados en la Tabla 2.

1.3.2 Análisis comparativo del promotor *BC1*

El análisis de la región corriente arriba del gen *BC1* no condujo a la identificación de CMA's exceptuando los que están presentes en la RC, que fueron previamente identificados al analizar el promotor del gen *AC1* incluido en esa región (Argüello-Astorga et al. 1994; Gomez-Castañón 2004). Si bien no fue posible identificar algún CMA en el promotor *BC1* fuera de la RC, el análisis hizo posible la identificación de una huella filogenética presente dentro de esa región no codificante en la mayoría de los BGVs analizados. Esta huella filogenética es un elemento con simetría de díada parcial que tiene una longitud de 13 nts (GGCCGCGCAGCGG), a la que denominamos "Secuencia Palindrómica Conservada" (Conserved Palindromic Sequence, CPS). La CPS está presente en casi todos los promotores *BC1* examinados, y no conserva su posición relativa dentro del promotor en todos los linajes. Su posición relativa al codón de inicio del gen *BC1* correlaciona de modo significativo con la región geográfica de la que son nativos los begomovirus (Fig. 10). Por ejemplo, se localiza a 3-4 pb corriente arriba del gen *BC1* en el subclado "India" de los BGVs del LP Asiático, a 11-13 pb arriba de *BC1* en los BGVs del continente americano y los Corchovirus, y a 125-160 pb de distancia en los BGVs de África. Consecuentemente, este elemento de secuencia puede considerarse una firma genómica biogeográfica, lo cual ha sido de utilidad para hacer algunas inferencias evolutivas que mencionaré más adelante.

Dado que la CPS está presente en la gran mayoría de los componentes B analizados, decidimos buscar ésta secuencia en el ADN-A correspondiente. La búsqueda mostró que la secuencia CPS está presente en todos los componentes A de

BGVs, y se localiza dentro del gen *AC1*, en la región que codifica los aminoácidos del 209 al 212 (AARP) de la proteína Rep (Fig. 11). También encontramos dicho elemento en los genomas de curtovirus y topocovirus pero no en los miembros del género *Mastrevirus*.

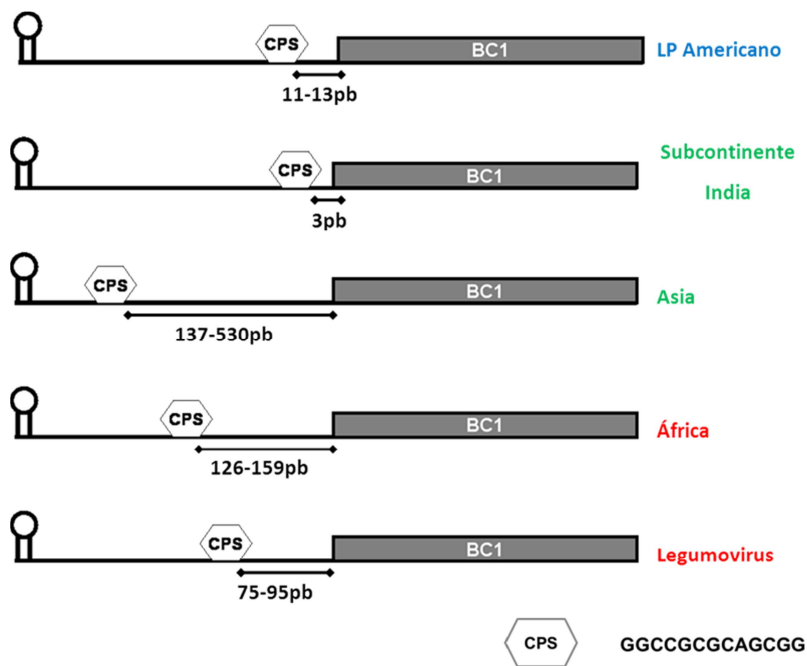


Figura 10. Ubicación de la CPS en el promotor de *BC1*

Ubicación de la secuencia palindrómica conservada en los promotores *BC1* de virus de los tres LP. La posición de la CPS, representada por un hexágono, varía en relación al origen geográfico del virus. Los números indican la distancia mínima y máxima que existe entre la CPS y el codón de inicio de la proteína *BC1*.

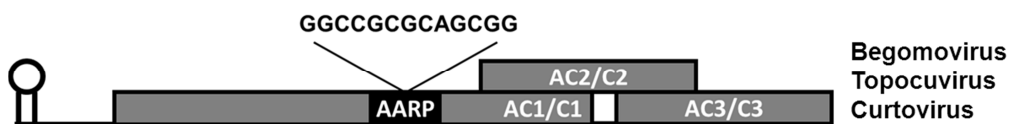


Figura 11. Localización de la CPS en el componente A.

La CPS se localiza en la región codificante del gen *AC1* de los géneros *Begomovirus*, *Curtovirus* y *Topocovirus*. El recuadro negro dentro del gen *AC1* indica los aminoácidos que son codificados en la secuencia de 13 nts (CPS).

La presencia de la CPS río arriba del gene *BC1* sugiere la posibilidad de que ese elemento palindrómico sea el sitio de unión de un factor transcripcional y, por lo tanto, funcione como un elemento cis-regulador de la transcripción. En el virus MYMV la expresión de los genes *AC2* y *AC3* ocurre a partir de un mismo transcrito, el cual se genera a partir de un promotor fuerte localizado arriba del gen *AC2* (Shivaprasad et al. 2005). Como se puede observar en la figura 11 la CPS se encuentra en la región 5' del gen *AC2* lo que es consistente con la suposición de que esa secuencia podría participar en el control de la expresión de *AC2*.

Por otra parte, se ha observado que trinucleótidos que se encuentran en el centro de secuencias palindrómicas favorecen la formación de estructuras tallo-asa compactas y estables (Kannan & Zacharias 2007); en particular, el trinucleótido GCA rodeado por las bases C y G (gcGCAgc) (Kannan & Zacharias 2011). La CPS es un elemento simétrico que tiene en el centro el trinucleótido GCA (GGCCGCGCAGCGG). Teniendo en mente la posibilidad de que la CPS forme una estructura tallo-asa, enviamos la secuencia al servidor RNAfold Web Server, el cual predice estructuras secundarias en ácidos nucleicos (Gruber et al. 2008). La predicción indicó que la CPS probablemente forme una estructura tallo-asa estable, y tendría, dadas sus características, la capacidad de formarse espontáneamente (-3.76kcal/mol) (Fig. 12).

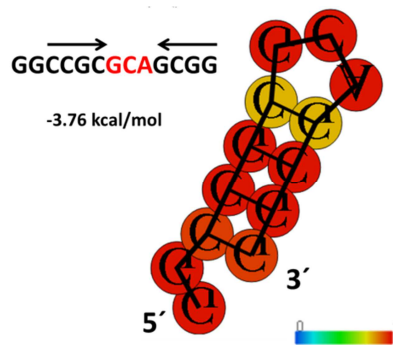


Figura 12. Predicción de la estructura secundaria de la CPS

La CPS está formada de un núcleo trinucleótido (letras en rojo) rodeado por secuencias palindrómicas (indicadas con flechas), la estructura tiene una energía libre de 3.76kcal/mol. La barra indica la probabilidad de que cada par de bases interaccionen.

El hecho de que la CPS posea la capacidad potencial de formar una estructura tallo-asa abre la posibilidad de que dicho elemento tenga una función distinta a la del control transcripcional de *BC1*. Por ejemplo, podría participar en la regulación a nivel traduccional, ya que la CPS se encuentra en la región 5' UTR de los transcritos que han sido mapeados en BGVs (Frischmuth et al. 1991; Sunter & Bisaro 1989; Shivaprasad et al. 2005). Se ha descrito que estructuras tallo-asa que se encuentran entre el 5'-cap y el codón de inicio pueden interferir con el ensamblaje adecuado del complejo de pre-iniciación de la traducción, o evitar el escaneo del transcrito por el ribosoma (Cazzola & Skoda 2000). También se ha reportado que las estructuras tallo-asa dentro de la región 5' UTR son necesarias para la estabilidad del ARNm y para una alta eficiencia de la traducción (Zou et al. 2003).

Otra función potencial de la CPS sería su participación en la síntesis de la cadena complementaria para formar el intermediario de doble cadena, el cual es necesario para la expresión de las proteínas virales. Se ha mostrado que las estructuras tallo-asa son importantes para la síntesis de la cadena complementaria, ya

que la estructura expone el sitio de reconocimiento de la primasa (Frick & Richardson 2001). Una observación interesante que es congruente con esta idea es el hecho de que los mastrevirus son los únicos que carecen de CPS y estos virus encapsidan un ADN de cadena sencilla corto, que funciona como “primer” para iniciar la síntesis de la cadena complementaria. Para el resto de los geminivirus no se ha descrito la presencia de un ADN cebador dentro del virión; y en cambio todos ellos poseen un elemento CPS. Otra observación importante es la posición de la CPS en el componente B, la cual difiere entre linajes biogeográficos de BGVs, como se ilustra en la Figura 10. Esta variabilidad en la localización de la CPS sugiere que la posición en el DNA-B no es importante para su función, lo que descartaría su función en la traducción, y apoyaría la posibilidad de que la CPS esté involucrada más bien en el control transcripcional de *BC1* o en la síntesis de la cadena complementaria.

I.4 – Identificación de la India como Centro de Origen de los begomovirus

La existencia de linajes continentales de BGVs indica que éstos evolucionaron en aislamiento geográfico durante períodos muy extensos. Por lo tanto, la dispersión inicial (“radiación”) de los BGVs debe haber ocurrido cuando algunos continentes estaban todavía unidos formando el super-continente Gondwana, hace aproximadamente 130Ma (A. G. Smith et al. 1994). En un reciente trabajo sobre la evolución de la familia *Geminiviridae*, Nawaz-ul-Rehman y Fauquet (2009) identificaron 8 centros de diversificación de geminivirus: A- Australia, B- Japón, C- Sur de China/SE Asiático, D- India, E- Africa subsahariana, F- la región del Mediterráneo y Europa, G- America del Sur, y H- Centroamérica. Sobre la base de la riqueza de especies virales en estas

regiones, concentradas principalmente en los centros C y D, estos autores concluyen que el Centro de Origen (CdeO) de los geminivirus muy probablemente se localiza en la amplia región Indo-China (incluyendo Pakistán, Bangladesh, Sri Lanka, Myanmar, Tailandia, Malasia, Camboya, Laos y Vietnam, aparte de la India y sur de China). En lo que respecta específicamente a los BGVs, algunos autores han postulado como su CdeO al Sureste Asiático, especialmente la región que comprende a Vietnam y los países vecinos, pero discrepan en la antigüedad atribuída al ancestro común de los BGVs. En efecto, Ha et al. (2008) consideran que el BGV ancestral surgió evolutivamente antes de la fragmentación de Gondwana (~130 Ma), en tanto que Lefeuvre et al. (2011) han estimado, por diferentes criterios y líneas de evidencia, que los BGVs aparecieron hace 30 a 70 Ma. Tomando en consideración la gran extensión de la región Indo-China postulada por Nawaz-ul-Rehman y Fauquet (2009) como el CdeO de los geminivirus, nos preguntamos si se podría circunscribir más estrechamente el centro originario de los BGVs utilizando los datos obtenidos en nuestros análisis, y tomando en cuenta la distribución y movimiento de los continentes en el Cretácico. El análisis combinado de los diversos datos moleculares (tales como las “firmas” en proteínas virales, la composición y organización de secuencias relevantes en el origen de replicación, el arreglo de CMAs en los promotores, etc.) junto con las filogenias derivadas de los componentes genómicos y las proteínas virales, así como la distribución geográfica de los LPs, nos condujo a la conclusión de que el área geográfica en la que más probablemente evolucionaron los BGVs es el subcontinente Indio (integrado por la India, Pakistan y Bangladesh). Algunas de las líneas de evidencia y argumentos que apoyan esa conclusión son las siguientes:

1) La reconstrucción de la arquitectura ancestral de la región promotora del gen *CP* por medio del análisis filogenético-estructural (Gomez-Castañón 2004) y datos de este estudio) permitió establecer que la organización primitiva de esa región de control se conserva todavía, en mayor o menor grado, en varias especies de BGVs aislados en la India, Tailandia, Vietnam y China (Fig. 13), pero no en virus de otras regiones del mundo.

2) El subcontinente Indio y Vietnam son las únicas regiones del mundo en las que coexisten BGVs perteneciente a los tres Linajes Primarios. En el caso de Vietnam, aunque posee una gran variedad de BGVs del LP Asiático y contiene además a las dos especies conocidas de Corchovirus (Ha et al. 2008; 2006), solo incluye a una especie de Legumovirus (*Kudzu mosaic virus*). En contraste, el subcontinente Indio posee la mayor riqueza de especies de Legumovirus del mundo y contiene también una enorme variedad de BGVs del LP Asiático (Bridson et al. 2010), además de incluir a una de las dos especies de Corchovirus descritas al presente.

3) La filogenia de los BGVs sugiere que su radiación inicial tuvo lugar en el período en que Gondwana empezó a fragmentarse, con el subsecuente aislamiento de estirpes virales en diferentes bloques continentales (Rybicki 1994). En este período de fragmentación y movimiento de los continentes (130-90 Ma) la India estaba separada y muy distante del resto de Asia. El contacto entre las placas tectónicas de la India y Asia se inicio hace unos 50 Ma cuando el subcontinente Indio colisionó con el continente asiático e inició la formación de la cordillera de los Himalayas (http://geosci.uchicago.edu/~rowley/Rowley/Collision_Age.html). Tomando en cuenta

estos eventos geológicos, la existencia de BGVs pertenecientes a los tres LPs en el subcontinente Indio podría explicarse de dos formas: a) los BGVs surgieron evolutivamente en la India y de ahí migraron a otros continentes, o b) los BGVs aparecieron primero en el SE asiático y migraron a la India tiempo después de la colisión de ésta con la placa tectónica Asiática. Esta segunda opción parece menos probable porque no explica la mayor riqueza de Legumovirus (ancestros de los BGVs de África) en la India, ni ofrece un escenario plausible para el establecimiento y diferenciación de los linajes biogeográficos de África y América, cuyas “firmas” moleculares se encuentran en varios BGVs de la India.

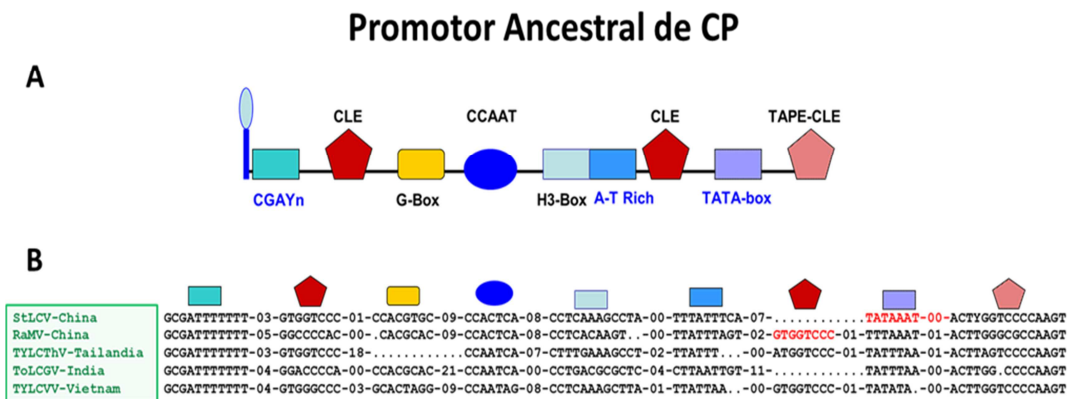


Figura 13. Estructura del promotor ancestral de CP.

El promotor ancestral de CP fue deducido en el trabajo de tesis de Gómez-Castañón (2004). A) Arquitectura del promotor ancestral. B) Promotores con la organización ancestral identificados en virus del Sur-Este Asiático y la India. Los elementos reguladores ilustrados son los siguientes: CGAYn=GCGATTTYYYY; CLE=GTGG(T/G)CCC; CCAAT=CAATCA; H3-Box=CCTCAAAGCCT; A-T Rich=ARWTAWWWW; TATA-Box=TATAWATA; TAPE-CLE=ACTTGGTCCCAAGT. Los números de acceso y los acrónimos están listados en la tabla 3.

I.5 – Organización genómica del Ancestro Común de los BGVs

El género *Begomovirus* incluye tanto virus con genomas monopartitas como bipartitas, y dado que el resto de los géneros de la familia *Geminiviridae* poseen genomas constituidos por una sola molécula de ADN, se tiende a suponer que los BGVs bipartitas se derivaron de un BGV ancestral monopartita. Sin embargo, diversas observaciones derivadas del presente trabajo ponen en cuestión esta suposición, y son las siguientes:

1) Dos de los tres LPs de BGVs incluyen tanto virus monopartitas como bipartitas, mientras que el LP que incluye a los BGVs de América y a los Corchovirus de Asia está formado exclusivamente por especies virales bipartitas (ver Fig.4 y 5). Considerando que los virus monopartitas de Asia y África son filogenéticamente distantes, y se asemejan más en secuencia nucleotídica global al componente genómico A de especies bipartitas pertenecientes a su propio Linaje Primario (v.gr: TYLCV y EACMV), sería necesario asumir, para mantener la validez de la hipótesis de que los BGVs bipartitas se derivaron de monopartitas, que el componente genómico B evolucionó independientemente en al menos tres ocasiones, una en África, otra en Asia, y una tercera en el continente americano, con la complicación adicional en este último caso de que no se ha descrito hasta ahora un BGV monopartita que pudiese considerarse el ancestro de los BGVs del Nuevo Mundo.

2) Los BGVs que presentan la organización ancestral de elementos *cis*-reguladores en el promotor *CP* (ver Fig. 13) son todos ellos bipartitas, con la excepción de *Tomato*

yellow leaf Vietnam virus (TYLCVV), el cual es altamente similar (88-87% de identidad global) al componente A de los virus bipartitas *Papaya leaf curl China virus* (PaLCCNV) y *Ageratum yellow vein China virus* (AYVCNV), respectivamente. TYLCVV se asocia a un beta-satélite, que parece substituir funcionalmente al ADN-B.

3) Las “firmas” en la proteína CP y las características del origen de replicación de los BGVs monopartitas son análogas a las observadas en los miembros bipartitas del Linaje Primario al que pertenecen.

4) Todos los virus que, por su posición en el árbol filogenético de los BGVs, pueden ser considerados los tipos ancestrales de cada uno de los tres Linajes Primarios y los clados secundarios (v.gr: Legumovirus, Corchovirus, virus americanos del Clado SLCV, etc.) poseen genomas bipartitas.

Por todas estas razones llegamos a la conclusión (algo inesperada) de que el ancestro común de todos los BGVs poseía, muy probablemente, un genoma compuesto por dos moléculas de ADN diferentes. Un importante corolario de esta hipótesis es que el proceso de evolución del segundo componente genómico probablemente comenzó mucho tiempo antes de que surgieran los geminivirus transmisibles por la mosquita blanca, y que los miembros más antiguos del linaje que dio origen a los BGVs pudiesen ser geminivirus transmitidos por otras especies de insectos relacionados a *Bemisia tabaci*, virus que no han sido todavía descubiertos.

Parte 2

Origen de los begomovirus del Nuevo Mundo

Los BGVs fueron clasificados inicialmente en dos grandes grupos, el del “Nuevo Mundo” y el del “Viejo Mundo”, sobre la base de los datos aportados por los primeros estudios filogenéticos de la familia *Geminiviridae*. Con el incremento en el número de especies de BGVs descritas, el grupo de BGVs del “Viejo Mundo” mostró ser muy heterogéneo, y actualmente se reconocen varios linajes del VM, como los de “Asia”, “India”, “Japón”, “África”, Legumovirus y Corchovirus (Briddon et al., 2010). Los BGVs del NM, siendo un grupo genéticamente menos heterogéneo, se divide en un menor número de linajes, algunos de ellos definidos por características genómicas específicas y no por su distribución geográfica, como el clado del SLCV (*Squash leaf curl virus*), cuyos miembros se encuentran diseminados desde el sur de EUA hasta Brasil y Argentina (Briddon et al., 2010).

Las notables diferencias de los BGVs del NM respecto a sus parientes de otros continentes (v.gr: la ausencia del gen AV2, la organización estructural de su origen de replicación, y otras características previamente mencionadas), junto con su menor heterogeneidad genómica, sugieren un origen más reciente que los BGVs del VM. La cuestión del origen de los BGVs americanos ha sido abordada por varios investigadores, y las hipótesis sobre ese punto son muy variadas, tanto en lo relativo al origen geográfico de los BGVs que se introdujeron inicialmente a América (el “linaje fundador”), como respecto al tiempo en el tuvo lugar dicho evento. En efecto, se ha postulado alternativamente que los BGVs del NM se derivaron de BGVs de Africa

(Rybicki, 1994), de Asia (Lefeuvre et al., 2011), o más específicamente, del SE asiático (Ha et al., 2006, 2008). En lo que respecta a la antigüedad estimada de los BGVs del NM, los valores propuestos son extremadamente variables, y van desde los 30 a 43 mil años de quienes postulan una introducción por parte de grupos humanos que migraron de Asia a América a través del Estrecho de Bering (Ha et al., 2006), hasta los 125 Ma, correspondientes al período en que se separaron los bloques continentales de América, India y Asia (Rybicki et al., 1994; Ha et al., 2008). Un reciente estudio que utilizó métodos de estimación más rigurosos, basados en las tasas de sustitución “lentas” y “rápidas” de los geminivirus, desacredita las hipótesis extremas antes mencionadas, al concluir que la antigüedad de los BGVs del NM no puede ser menor a 2 millones de años ni mayor a 80 Ma, y sugieren que el período más probable en que los BGVs arribaron a América fue entre 20 y 30 Ma atrás, cuando existían condiciones climáticas propicias para la migración de la mosquita blanca a través del puente terrestre de Bering (Lefeuvre et al., 2011).

Si los BGVs no fueron introducidos al continente americano por el hombre, como se sugiere, resulta obvio suponer que llegaron a este continente asociados a poblaciones migrantes de su insecto vector, por lo que el hallazgo de fósiles de la mosquita blanca (v.gr: incrustaciones en ámbar) podría considerarse una evidencia indirecta de la presencia de BGVs en América, lo que permitiría estimar de modo más apropiado la edad mínima de los mismos. Desafortunadamente, no se han encontrado al presente fósiles de *Bemisia tabaci* en el continente americano. Existen, sin embargo, otro tipo de fósiles que proporcionan información sobre la antigüedad de los BGVs del NM. En efecto, en 1996 se descubrió una serie de secuencias repetidas con similitud al

ADN-A de BGVs, a las que se denominó “elementos GRD” (*Geminivirus-Related DNA*), los cuales se encuentran integrados en el genoma de varias especies del género *Nicotiana* (Bejarano et al., 1996). El análisis detallado de los GRDs presentes en 5 especies de este género (*N. tabacum*, *N. tomentosiformis*, *N. tomentosa*, *N. kawakami*, y *N. otophora*) llevó a distinguir dos clases de GRDs (GRD3 y GRD5), que corresponden a dos eventos de integración independientes, temporalmente separados (Murad et al., 2004). El tiempo en el que ocurrieron esos eventos ha sido estimado por diferentes métodos, y los valores obtenidos para la primera integración van de un mínimo de 2 Ma a poco más de 7 Ma (Murad et al., 2004; Gibbs et al., 2010; Lefeuvre et al., 2011). Este marco temporal es importante para elaborar hipótesis plausibles sobre el origen de los BGVs del NM, y en particular, para establecer criterios que permitan descartar rutas de migración de los BGVs y su vector hacia América, por su inviabilidad en el contexto de la posición y desplazamiento de las masas continentales en el período considerado.

Rutas potenciales de migración al NM

Se pueden proponer varias rutas de migración de los BGVs desde su Centro de Origen hasta el continente americano, dependiendo de la posición relativa de esas dos regiones geográficas al tiempo en que comenzó la colonización de América por mosquitos blancos virulíferos: 1) Una posibilidad es que los BGVs hayan pasado del centro de origen (India) a África y posteriormente a América, ya que el continente Africano se separó de América en la última fase de la fragmentación del supercontinente. Gondwana; 2) Otra ruta migratoria posible iría de Asia a América a

través de lo que hoy es el estrecho de Bering. Se ha propuesto que esta migración pudo ocurrir hace 30 Ma, cuando el puente terrestre de Bering era de carácter subtropical (Lefeuvre et al. 2011); 3) Una ruta alternativa pero rara vez considerada es la que involucra al continente Antártico, que fue el bloque continental de Gondwana que mantuvo conexiones terrestres con Sudamérica hasta una época más reciente, cerca de 30 Ma (<http://en.wikipedia.org/wiki/Gondwana>). Analizaremos en lo que sigue esas rutas alternativas a la luz de la información existente, incluida la obtenida en el presente trabajo.

La ruta de migración “Africana”

África y Sudamérica se mantuvieron unidas por puentes terrestres hasta hace unos 100 Ma, en el período Albiano (Smith et al., 1994). En ese tiempo la India (hipotético CdeO de los BGVs) se encontraba muy cerca de Africa pero no tenía conexión terrestre con ese continente ni ningún otro bloque continental generado por la fragmentación de Gondwana. Así pues, parece improbable que los BGVs hayan llegado a Sudamérica vía Africa, lo que implicaría además que los BGVs del NM tendrían una antigüedad igual o mayor a 100 Ma, lo que no es consistente con los datos de divergencia genómica entre estos virus, que son menos heterogéneos que los BGVs del VM, considerablemente más antiguos.

Al margen de las cuestiones paleobiogeográficas, hay datos generados en el presente estudio que prácticamente excluyen la posibilidad de que el o los linajes fundadores de los BGVs del NM provengan de BGVs de África. Por ejemplo, de

acuerdo a nuestro análisis, más del 90% de los BGVs del continente africano y Madagascar exhiben el motivo GGPSGM en el segmento 189-194 de su proteína CP, motivo que no se encuentra en la proteína homóloga de ningún BGV del NM, que en vez de ello exhiben la secuencia GGQYAS/C, común a los Legumovirus y a los BGVs del LP Asiático. De modo análogo, la ausencia de motivos comunes a las proteínas BC1 y BV1 de los virus Africanos y Americanos, con exclusión de otros linajes, así como las marcadas diferencias en la organización estructural del origen de replicación de estos dos grupos virales, indican que los BGVs americanos no provienen de virus de Africa.

La ruta de migración de Asia a América por el estrecho de Bering

La mayor parte de las hipótesis modernas sobre el origen de los BGVs del NM asumen que los ancestros de los mismos fueron de origen asiático, particularmente a raíz del descubrimiento de los Corchovirus de Vietnam y la India (Ha et al., 2006, 2008) que están claramente relacionados con los virus americanos, pero que son suficientemente divergentes en secuencia para excluir la posibilidad de que se trate de BGVs americanos introducidos por el hombre al SE Asiático. Dado que en la configuración geográfica actual la vía más plausible para el tránsito de especies biológicas entre Asia y América es el estrecho de Bering, las hipótesis convergen en este punto y solo se diferencian en lo concerniente al período en que la migración de las mosquitas blancas virulíferas pudo haber ocurrido, como se ha mencionado previamente. Hay sin embargo un serio problema con este escenario, y es la inevitable conclusión de que los primeros BGVs evolucionaron en Norteamérica y de ahí habrían radiado al resto del continente

americano. Sin embargo, esta dispersión pan-americana solo habría podido haber ocurrido después de que la conexión entre el bloque Norteamericano y Sudamérica se consolidó, al formarse el Istmo de Panama, evento que tuvo lugar hace tan solo 3 Ma. Esto no concuerda con la edad de los fósiles moleculares de begomovirus (los GRDs) integrados en el genoma de plantas del género *Nicotiana*, que se ha estimado en 6 ó 7.5 Ma (Lefeuvre et al. 2011; Murad et al. 2004; Okamuro & Goldberg 1985). Más aún, las especies de *Nicotiana* que contienen secuencias GRD en su genoma son todas de origen Sudamericano, y por lo tanto fueron infectadas por BGVs presentes en un subcontinente (América del Sur) que estuvo separado de Norteamérica a lo largo de un período de tiempo de más de 150 Ma antes de la formación del Istmo de Panamá (Smith et al., 1994)

Dado que la migración de los BGVs al continente americano no parece haber ocurrido a través de África o por el estrecho de Bering, nosotros proponemos que los begomovirus de América se derivaron de un linaje que migró de la India al continente Antártico, y de ahí a Sudamérica, en una época en que todavía existían puentes terrestres entre estos dos continentes. Según la migración que proponemos, la Antártida fue usada por los begomovirus como un puente entre India y Sur-América. Este puente continental fue usado por especies del género *Nicotiana* (Aoki & Ito 2000) y por marsupiales (Woodburne & Case 1996) para la migración entre Sur-América y Australia. Los principales datos que apoyan esta hipótesis son los que obtuvimos del análisis de las regiones promotoras de *CP* y *BC1*, y de las regiones codificantes (motivos de *BV1*, *BC1* y *CP*), que relacionan los virus americanos con los asiáticos.

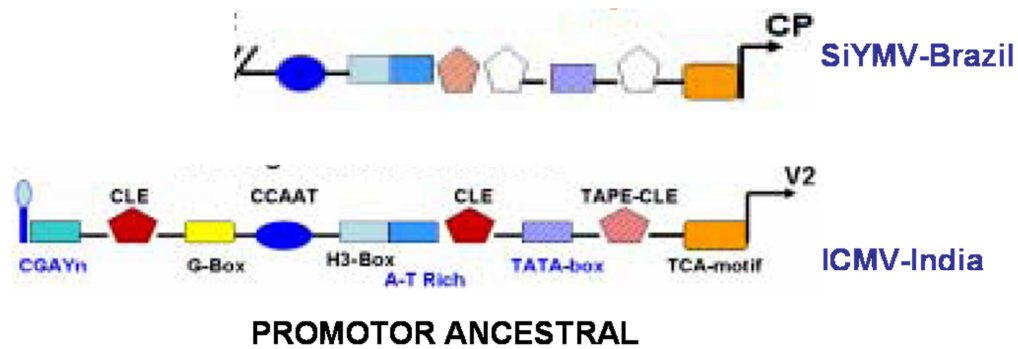


Figura14. Similitudes entre el promotor Ancestral de CP y el promotor CP de virus aislados en Sur-América.

El promotor del virus *Sida yellow mosaic virus* (*SiYMV*) conserva la secuencia y distancia de seis posibles sitios de unión de factores de transcripción (CCAAT-Box; H3-Box; A-T Rich; CLE; TATA-Box; Motif TCA) presentes en el promotor ancestral de CP, que está presente en virus de la India. Figura modificada de (Gomez-Castañón 2004).

Gómez-Castañón (2004) encontró que algunos promotores CP de virus de Sur-América, v.gr. el virus *Sida yellow mosaic virus*, comparten potenciales sitios de unión de factores de transcripción que se conservan en secuencia y en posición con el promotor ancestral CP (Fig. 14). Gómez-Castañón y colaboradores (2004) encontraron esta similitud cuando compararon promotores los CP de virus aislados en el subcontinente Indio y en sus alrededores, que aún conservan la estructura ancestral, con virus aislados de otras regiones geográficas con el fin de trazar la evolución del promotor. Dentro de la región del promotor del gen de BC1 nosotros localizamos la secuencia CPS. Esta secuencia es indicativa de la región geográfica a la que pertenece cada virus. Como se observa en la figura 10, en los virus originarios del Sub-continente de la India la CPS se localiza a 3 nucleótidos del ATG de BC1, distancia similar a la que presentan los del LP Americano/Corchovirus (12 nucleótidos). Sin duda esto también

relaciona a los virus americanos con los virus de la India y apoya nuestra propuesta de la migración desde la India hasta América.

Las evidencias que encontramos en las proteínas de begomovirus que apoyan nuestra hipótesis son: 1) el motivo que comprende los aminoácidos 190-195 de la proteína CP la cual relaciona a los virus americanos y asiáticos (Fig. 8B) y 2) las firmas presentes en los motivos de las proteínas BC1 y BV1 que relacionan a los virus de Vietnam (Corchovirus) con los virus Americanos y confirman que el LP Americano/Corchovirus es real. Otra evidencia que apoya nuestra propuesta es la gran diversidad de BGVs que existe en Sudamérica. En Sudamérica encontramos begomovirus atípicos del clado SLCV que tienen una organización de la RI diferente al resto de los virus americanos. También en esta región geográfica se han descrito dos satélites alfa en Brasil (Paprotka et al. 2010) y uno en Venezuela (Romay et al. 2010). Dado que en el centro en los centros de origen (India) es la región de mayor riqueza es probable que las regiones geográficas más cercanas al origen también lo sean. Por lo tanto la gran diversidad que existe en el Sur-América nos indica que la ruta de migración propuesta por nosotros es muy probable.

Conclusión, centro de origen de los begomovirus y migración al NM

El conjunto de datos obtenidos de la comparación de secuencias mediante el análisis pareado de las proteínas y la búsqueda de CMAs en los promotores de los begomovirus nos permitió establecer aspectos importantes de la evolución de estos virus. La reconstrucción filogenética de los componentes A y B obtenida del análisis

bayesiano confirmó lo previamente establecido, que ambos componentes evolucionaron en aislamiento geográfico estricto, los virus se separan en clados muy bien definidos de acuerdo a la región geográfica de donde fueron aislados. El hecho de que el análisis filogenético muestre este patrón es indicativo de que los componentes genómicos A y B estuvieron presentes antes de la separación de los continentes y que estos comparten la misma historia evolutiva. Analizando la región codificante y promotora de los componentes A y B, establecimos que muy probablemente el origen de los begomovirus es el sub continente de la India. Esta conclusión fue concebida debido a que las firmas geográficas de los LPs presentes en las proteínas BV1 y BC1 siempre son compartidas con virus de esta región geográfica, además de que el promotor ancestral de CP se encuentra presente en virus del sub-continente de la India y en virus asiáticos. Integrando los datos sobre movimiento de los continentes y de las condiciones climáticas que imperaban en el súper-continente Gondwana, con los datos obtenidos de la comparación genómica, establecimos que el origen de los begomovirus Americanos no puede ser a partir de los virus de África, pero sí a partir de los asiáticos. Con estas evidencias formulamos la hipótesis de la migración de los begomovirus asiáticos a América, en la que postulamos que la ruta más probable incluye el continente antártico. El continente antártico pudo funcionar como puente migratorio de los begomovirus entre el centro de origen y el continente americano.

Parte 3

Evolución del segundo componente de los Begomovirus

Integrando el análisis comparativo de las regiones codificantes y no-codificantes de los componentes genómicos A y B, encontramos evidencias de que el promotor y la proteína BV1 son verdaderos homólogos al promotor y proteína de CP, presentes en el genoma A. Además proponemos que el promotor y la proteína BC1 surgieron del arreglo genómico de la región complementaria de un componente A de begomovirus. Nuestros datos sugieren fuertemente que el ADN-B de los begomovirus se derivó evolutivamente de un ADN-A.

Las proteínas BV1 y CP son homólogas

Pensando en la idea de un origen evolutivo del componente B que no se relacione al componente A de los geminivirus, supusimos que las proteínas del ADN-B pudieran provenir de las plantas, de un satélite o de otros virus. Para explorar esta idea decidimos realizar búsquedas de la proteína BV1 contra la base de datos NCBI, excluyendo las proteínas de geminivirus. Para realizar la búsqueda utilizamos la herramienta "Position-Specific Iterated BLAST" (PSI-BLAST), la cual es un variante de BLAST útil para encontrar proteínas muy distantes que están relacionadas (Altschul et al. 1997). Las secuencias de aminoácidos de BV1 que utilizamos son de virus representativos de los tres linajes primarios los cuales se encuentran enlistados en la tabla 3. La búsqueda de homólogos de la proteína BV1 realizada con PSI-BLAST, arrojó que las proteínas BV1 no están relacionadas con proteínas de movimiento de virus o plantas, lo que sugiere que BV1 no fue adquirida por recombinación de un virus de plantas o de plantas asiladas en la actualidad (Datos no mostrados).

Otra alternativa para explicar el origen de la proteína BV1 es la idea propuesta por Kikuno y colaboradores (1984), que sugiere que la proteína BV1 es homóloga a CP y que se originó de la modificación del gen CP del ADN-A. Para explorar esta idea decidimos repetir la búsqueda con PSI-BLAST, esta vez el análisis fue dirigido contra las proteínas CP de geminivirus. En este caso, los resultados de las búsquedas para cada una de las proteínas BV1 de los virus de los tres linajes primarios arrojaron indicios de la homología entre las proteínas BV1 y CP de geminivirus. Varias proteínas BV1 comparten bloques de aminoácidos con la proteína CP; sin embargo, algunos bloques conservados entre CP y BV1 de una especie son diferentes a los bloques entre CP y BV1 de otra especie. Con los bloques que identificados en diferentes comparaciones, buscamos una proteína BV1 que en su secuencia fueran representados el mayor número de bloques y encontramos que la proteína BV1 del virus *Clerodendrum golden mosaic China virus* (CIGMCNV) es la que mayor número de bloques comparte con las proteínas de CP. En esta proteína nosotros definimos ocho bloques de aminoácidos distribuidos a lo largo del alineamiento entre las proteínas CP de begomovirus de los tres LPs y BV1 de CIGMCNV (Fig. 15). Los bloques se definieron de acuerdo a los siguientes parámetros: 1) La similitud promedio entre los aminoácidos de cada bloque debe de ser mayor al 50%; 2) El número de aminoácidos que comprenden los bloques deben de ser de 10 o más; 3) El espaciamiento entre residuos conservados no debe de superar los 5 aminoácidos; 4) Sólo un gap con una extensión de uno es permitido dentro del bloque.

Revisamos en la literatura si existen reportadas funciones para los bloques conservados que encontramos entre BV1 y CP. No encontramos que los aminoácidos

que forman los bloques conservados hayan sido correlacionados con alguna función específica de las proteínas BV1 y CP, excepto por algunos aminoácidos que forman los bloques I y III. Los bloques I y III se encuentran dentro de la región N-terminal de las proteínas NSP y CP. Ésta región es un dominio básico (rico en K,R) conservado. Dentro de la región N-terminal entre los aminoácido 3-20 de CP encontramos que existe una señal bipartita de localización nuclear (SLN) funcional, la cual fue descrita para el virus TYLCV (3 KR-X11-KvRRR 20) (Kunik et al. 1998), que está conservada en la CP de los BGVs (Sharma & Ikegami 2009). Como se observa en la figura 15, dentro del bloque I se localiza una parte de la SLN bipartita (KvRRR), las tres últimas argininas están presentes en la proteína BV1 de CIGMCNV. Analizando las secuencias de BV1 encontramos que algunos aminoácidos que forman la SLN en las proteínas CP están presentes en BV1 pero nunca una SLN completa como la que se encuentra presente en CP. Al revisar la literatura encontramos que las SLN que se han reportado para BV1 también caen el N-terminal. Se han identificado 2 SLN funcionales, en BV1 de BDMV se identificaron entre los aminoácidos 27 al 39 (27R[L/T]S[A/I][N/V]KRHDGKRR39) y entre los aminoácidos 81-96 (81[S/Q][K/L]GKMEPNR[S/C]R96) (Zhou et al. 2007). En SLCV se identificó una SLN bipartita entre los aminoácidos 25 al 39 (25 KR-X9-RRRP 39) y una SLN monopartita entre los aminoácidos 87al 95 (87PNRTRTYIK95) (Sanderfoot et al. 1996). Aunque están presentes las SLN en BV1, éstas no se localizan en la misma posición que en CP.

En el dominio básico de las proteínas CP y BV1se ha localizado el sitio de unión a DNA. En CP de MSV y en BV1 de SLCV se ha descrito que el sitio de unión al DNA se encuentra en la región N-terminal (Liu et al. 1997; Pascal et al. 1994). Tanto la CP

como la BV1 de estos virus, no presentan un consenso en el que se distinga el sitio de unión pero las mutaciones en esta región eliminan la unión al DNA. Caso contrario es el sitio de unión al DNA de la proteína CP de BGVs (*Tomato leaf curl Bangalore virus*) el cual es un motivo dedo de zinc (54-114aa) del tipo C2H2, que se encuentra conservado entre los begomovirus de los tres linajes principales (Kirthi & Savithri 2003). El bloque III se localiza entre los aminoácidos 65 al 85 de CP, región en la cual se ha mapeado un dedo de zinc. Dentro de este bloque hay dos cisteínas responsables de la formación del dedo de zinc, los cuales se substituyen por serinas en BV1 de CIGMCNV.

Los bloques encontrados no se correlacionan con regiones conservadas de la proteína BV1, excepto por el bloque VII. La mayoría de los aminoácidos que constituyen este bloque están conservados en todas las proteínas BV1 y CP analizadas (residuos resaltados en azul) (Fig. 15), en especial la región que contiene los aminoácidos ENALLLY. Esta región es rica en leucina, requisito necesario para las señales de exportación nuclear (SEN). El software “NetNES 1.1 Server” que predice SEN reveló que la secuencia conservada del bloque VII posiblemente sea una SEN (la Cour et al. 2004).

La existencia de ocho bloques conservados sugiere fuertemente que las proteínas CP y BV1 son homólogas, sumado a estos bloques existen características compartidas entre estas proteínas que apoyan esta hipótesis. La principal observación es el hecho de que las proteínas CP funcionen como como proteínas lanzadera (algunas proteínas CP de virus del NM, todas las CP de BGVs monopartitas y la CP de mastrevirus). De hecho, la CP del virus SLCV tiene la capacidad de complementar mutantes de BV1 (Ingham, et al. 1995). Otra observación es que los genes y las proteínas tienen la misma posición relativa y longitud. La comparación global de las proteínas CP y BV1 del virus ACMV muestra que estas proteínas tienen una similitud del 43% (Kikuno et al. 1984).

Los promotores de BV1 y CP son homólogos

Basados en el resultado anterior, nosotros decidimos comparar las regiones de los promotores de los genes *BV1* y *CP* con el objetivo de buscar elementos reguladores compartidos por ambos promotores y verificar si estos promotores son homólogos. Los promotores *BV1* de especies de los diferentes LP los comparamos contra la base de datos de nucleótidos del componente A de geminivirus del NCBI, utilizando la herramienta para nucleótidos BLASTn. El mismo análisis lo realizamos con los promotores *CP*, en este caso se compararon los promotores contra la base de datos de nucleótidos del componente B. La región de los promotores la delimitamos entre el primer nucleótido del virus (sitio de corte) hasta el ATG de la proteína *CP* o *BV1*. El análisis arrojó numerosos ejemplos de la similitud que existe entre los promotores *CP* y *BV1*. En algunos begomovirus encontramos que el promotor de *BV1* tiene una extensa similitud con el promotor *CP*, tanto en el arreglo como en el espaciamiento de secuencias reguladoras. Por ejemplo el promotor *BV1* de *Bean calico mosaic virus* conserva elementos del promotor *CP* de los virus *Tobacco yellow crinkle virus*, *Tomato yellow mottle virus* y *Pepper golden mosaic virus* (Fig.16).

	CLE	CAAT-BOX
BCaMV	GGgGTCCC-06-AATTTGAA-05-CCAATCA-20-GATATTT-02-GACTTG-05-AAGTTGTGG	
TbYCrV	*GTGGTCCC-11-AATTTGAA-05-CCAATCA-20-GATATTa-02-GACTTG-05-AAGTTGTGG	
ToYMoV	*GTGGTCCC-17-.....CCAATCA-20-GATATca-02-GACTTG-03-AAGTTGTGG	
PepGMV	*GTGGTCCC-04-TATTTGAt-09-CCAATCA-20-GATATTT-14-.....AAGgTGTGG	

Figura 16. Comparación del promotor de BV1 contra el promotor de CP
 La figura muestra el alineamiento entre el promotor BV1 de *Bean calico mosaic virus* (resaltado en amarillo) en contra de los promotores CP de tres especies de virus americanos. En esta comparación se identificaron elementos tardíos conservados (CLE) y cajas CAAT. * elemento regulatorio en la cadena negativa. Los números de acceso y acrónimos se encuentran en la tabla 3.

El ejemplo más claro que ilustra la similitud entre los promotores CP y BV1 es la comparación del promotor CP de ToMiMV. Este promotor es similar a los promotores BV1 de especies aisladas en Sur-América (ToMiMV:EU710753, SiMiMV:HM545834, SiMBoV:HM585442 y OkMoV:EU914818). En particular el análisis mostró que el promotor completo de CP del virus ToMiMV se encuentra embebido en el promotor de BV1 SiMBoV, y que estos son 69 % idénticos. En estos promotores nosotros identificamos ocho bloques compartidos, los cuales se encuentran distribuidos a lo largo del alineamiento (Fig. 17). Dentro de los bloques localizamos elementos reguladores que se conservan en secuencia y en el arreglo modular. La identidad de los elementos reguladores conservados se obtuvo de una búsqueda realizada en la base de datos PlantCARE (Thijs et al. 2002), entre los elementos identificados se encuentran cajas G, cajas CAAT y cajas W, entre otras (Fig. 17).

Adicionalmente se ha reportado que el promotor BV1 de *Euphorbia mosiac virus*-Jalisco es semejante al promotor CP del virus *Rhynchosia golden mosaic Yucatan virus* (Gregorio-Jorge et al. 2010). Sumado a la similitud en la estructura de estos promotores, se ha reportado que ambos son de expresión tardía, requieren de la transactivación de la proteína TrAP y conservan el elemento CLE (blanco de TrAP) (Sunter & Bisaro 1992; Ruiz-Medrano et al. 1999; Argüello-Astorga et al. 1994). Tomando en cuenta la similitud en secuencia que existe entre las regiones no codificantes río arriba de los genes CP y BV1 que analizamos, y las propiedades que comparten ambos promotores, nosotros sugerimos que el promotorBV1 posiblemente se derivo evolutivamente de un promotor CP de begomovirus.

Origen del Gen BV1

La existencia de ocho bloques conservados a lo largo del alineamiento entre proteínas CP y BV1 de CIGMCNV, las características compartidas entre estas proteínas y las similitudes en estructura y propiedades que existen entre los promotores BV1 y CP sugieren que el gen de BV1 es homólogo y se originó a partir del gen CP. Se ha observado que los transcritos de BV1 y de CP tienen el mismo tamaño aproximado, y en ambos casos el inicio de la transcripción se encuentra a 30 nts de la caja TATA (G. Sunter & Bisaro 1989). El conjunto de datos presentados aquí nos permiten concluir de manera inequívoca que los genes CP y BV1 son homólogos.

La proteína BC1 se originó de los genes sobrepuestos AC1, AC2 y AC3

Las proteínas del movimiento de begomovirus (BC1) poseen las mismas características de la familia de proteínas del movimiento (MP) de los virus de RNA, son capaces de unirse al DNA (Rojas et al. 1998), interacciona con los plasmodesmos (Lazarowitz & Beachy 1999), y virus como *Squash leaf curl virus* y se asocian con túbulos del retículo endoplásmico (Ward et al. 1997). Las características compartidas de las proteínas de la familia MP y BC1 sugieren que están relacionadas de alguna manera, por lo que decidimos buscar los homólogos de BC1 entre las proteínas MP. La búsqueda de los homólogos de BC1 la realizamos usando la herramienta BLASTP, las proteínas BC1 que utilizamos pertenecen a virus de los tres LP. El análisis de BLASTP lo dirigimos contra las proteínas MP de la base de datos del NCBI y excluimos las proteínas BC1. Este análisis mostró que las proteínas BC1 tienen una región de 45aa entre los aminoácido 48 al 95 que tiene una similitud del 57-64% con una región de aminoácidos

de las MP de virus de RNA que infectan plantas, en particular con las proteínas MP de Tobamovirus (Fig. 18), se sabe que esta región es la más conservada entre las proteínas del movimiento de la superfamilia de 30kda y es usada para identificar las proteínas que pertenecen a esta familia (Mushegian & Koonin 1993). Dentro de esta región encontramos un bloque conservado de cuatro aminoácidos idénticos consecutivos (DKRM o DTRM) entre algunas proteínas MP y BC1, el bloque contiene el motivo conservado de la familia de proteínas MP (DXR) el cual se cree es importante para la función de la proteína (Información consultada en junio del 2012 en la base de datos PFAM versión 26.0). (Fig. 18). El resultado anterior podría indicarnos que la proteína BC1 tiene su origen en las proteínas MP. Sin embargo, la región de similitud sólo representa el 15% de la longitud de BC1. Como esta región no cubre toda la proteína BC1 nosotros decidimos repetir el análisis de BLASTP y PSI BLAST, ahora dirigiendo la búsqueda contra las proteínas del componente A que ocupan la misma posición en el genoma en el sentido complementario (AC1, AC2 y AC3).

BC1

RhYMIV. 49 CMKLDH**V**VI**E**YRNQ**V**FNATGTVV**V**TI**D**K**R**ICDD-**E**AAQ**A**AFT**F**P 93
EACMCV. 48 CMKLDH**I**I**I**EYRNQ**V**FNATG**S**VI**V**EL**R**DN**R**V**S**LE-**D**AAQ**A**AFT**F**P 92
CLGMCNV 50 CLR**I**DH**V**I**E**YRNQ**V**FNATG**H**V**V**IE**M**H**D**TR**L**LEG-**D**S**R**Q**A**EFT**F**P 94
ToLCNDV 51 CMKLDH**V**V**L**EYRNQ**V**FN**A**VG**H**V**V**IE**L**H**D**M**R**L**T**EG-**D**T**R**Q**A**EFT**F**P 95
AbMV 50 CMKLDH**V**I**E**YR**Q**Q**V**FN**A**SG**T**V**V**VE**L**H**D**K**R**M**D**DN-**E**SL**Q**A**T**W**T**F**P** 94
CoGMV 50 CMKLDH**V**I**E**YR**Q**Q**V**FN**A**SG**S**V**I**VE**L**H**D**TR**M**DN-**E**SL**Q**T**T**Y**T**F**P** 94

MP de virus de ARN

CuGrMoMV 64 Y**V**T**L**LG**V**V**V**SGV**W**N**V**PE**S**CR**G**GAT**V**AL**V**D**T**RM**S**V**A**E**G**T**I**C**K**F**S**A**P** 109
CuMoV 69 Y**V**T**L**LG**V**V**I**SGV**W**T**I**PE**G**C**S**G**G**AT**V**GL**V**D**T**RM**S**ERV**V**E**G**T**V**C**K**F**S**V**P** 114
KyGrMoV 69 Y**V**T**L**VG**V**V**I**SGV**W**T**I**PE**N**C**A**G**G**AT**V**AL**V**D**T**RM**S**M**V**D**E**G**T**I**C**K**F**S**V**A 114
MaMV 78 Y**M**L**L**A**C**V**I**SG**R**W**H**V**E**NT**V**K**G**E**V**F**L**S**L**M**D**K**R**I**D**E**F**E**A**T**I**L**T**ANAN 123
PaFMV 78 Y**M**F**L**A**C**V**V**SG**R**W**H**V**E**PT**V**K**G**E**V**F**L**S**L**M**D**R**R**L**T**N**E**F**E**S**V**I**L**TANAK 123
BPepMoV 74 Y**I**C**L**AG**L**V**V**PG**E**W**N**L**P**DN**C**T**G**V**S**V**C**L**V**D**K**R**M**ER**A**S**E**AT**L**G**S**Y**T**T**G** 119
TobMGrMV 66 Y**V**C**L**AG**L**V**V**SG**E**W**N**L**P**DN**C**R**G**S**V**C**I**V**D**K**R**M**K**R**S**K**E**AT**L**G**A**Y**H**A**P** 111
HiLFPV 72 Y**V**H**L**LA**V**V**I**SG**K**W**H**L**H**NT**V**G**G**AT**I**AL**I**D**K**R**I**H**D**PK**Q**A**L**A**K**M**V**V**A** 117

Figura 18. Comparación entre proteínas BC1 y proteínas MP de *Tobamovirus*.

La búsqueda se realizó con la herramienta PSI-BLAST y el alineamiento entre las proteínas BC1 de virus de los tres LPs y las proteínas MP de los *Tobamovirus* se realizó de forma manual. Los aminoácidos idénticos o similares están resaltados en color amarillo, los aminoácidos resaltados en morado indican un segundo aminoácido conservado en una misma posición del alineamiento. Los residuos idénticos están en negritas y los similares (Blosum 62) se presentan en letras normales. Las cajas resaltan el bloque conservado de cuatro aminoácidos continuos conservados. Los residuos en color rojo, son los aminoácidos que conforman el motivo DXR característico de las

Análisis comparativo entre AC3 y la C-terminal de BC1

Las primeras comparaciones que realizamos entre las proteínas codificadas en el sentido complementario del ADN-A y BC1 arrojaron que la región C-terminal de BC1 tiene bloques conservados con proteínas AC3 de begomovirus de los tres LPs. Este resultado preliminar nos condujo a realizar comparaciones de forma dirigida entre la región C-terminal de BC1 y las proteínas AC3. Esta estrategia nos permitió encontrar cinco motivos conservados entre la proteína AC3 y la región C-terminal de BC1 (Fig. 19). Los cinco motivos están distribuidos en toda la proteína AC3 y entre los aminoácidos 150 hasta el final de la proteína BC1.

El motivo I está presente en las proteínas AC3 y BC1 de los tres LPs, este motivo tiene una longitud de 34 aa. En AC3 el motivo se extiende entre los aminoácidos 1-34 y en BC1 desde el aminoácido 150 hasta el 183. A lo largo del motivo de 34 aa encontramos que en 18 posiciones del alineamiento existen residuos idénticos entre BC1 y AC3 (Las posiciones con residuos idénticos entre alguna proteína de BC1 y alguna proteína AC3 están resaltadas en amarillo y la letra en negritas). Dentro de este motivo encontramos un bloque de cinco aminoácidos (WTVPN) de los cuales cuatro son idénticos entre los bloques de la proteína BC1 de PepYLCIV y las proteínas AC3 de MChMV y JaYMIV (Fig. 19a). Los bloques están formados por cuatro o más aminoácidos idénticos consecutivos, dentro de este bloque es permitido un aminoácido similar

El motivo II se encuentra entre los aminoácidos 27 y 42 de las proteínas AC3 de virus africanos y en BC1 entre los aminoácidos 184 y 199 de virus africanos y asiáticos. Este motivo tiene una longitud de 16 aminoácidos, en 10 posiciones del alineamiento existen residuos idénticos. Es interesante resaltar que en la proteína AC3 del virus TYLCV, el bloque de siete aminoácidos continuos TAHSQRP es idéntico al bloque presente en la proteína BC1 de los virus EACMZV y EACMCV (Fig. 19b).

El motivo III conformado de proteínas AC3 de los tres LPs y proteínas BC1 de virus americanos tiene una longitud de 32 aminoácidos, el motivo presenta 15 posiciones de residuos idénticos entre las proteínas comparadas. Dentro de este motivo existen dos bloques, uno compartido entre la proteína AC3 del virus asiático EmYVV y

cuatro virus americanos (IEIR) y el otro bloque compartido (ALGM) entre las proteínas BC1 de tres virus americano y las proteínas AC3 (Fig. 19c).

El motivo IV es interesante puesto que está presente en las proteínas AC3 y BC1 de los tres LPs. Este motivo es pequeño (16 aa) pero muy conservado (11 posiciones con residuos idénticos) y dentro de este se encuentra un bloque de cinco residuos, todos ellos son idénticos si comparamos la proteína BC1 de CICrV y las proteínas AC3 de los tres LPs (Fig. 19D).

El motivo V se encuentra en la región C-terminal de las dos proteínas. En AC3, el motivo llega hasta al final de la proteína y en BC1 el final del motivo se encuentra a seis aminoácidos del final de la proteína BC1. Dentro del motivo se encuentra un bloque de cuatro aminoácidos (IETT) presente en la proteína AC3 del virus *TYLCGuV* y en la proteína BC1 de los virus del LP Africanos/Legumovirus (Fig. 19e).

A**Motivo I - AC3/BC1****AC3**

MChMV 1 MDSRTGESITVYQAENSVFIWTVPNPLFFKIIYHV 34
BCaMV 1 MDSRTGEFITAAQAENSVFIWEVPNPLYFKIYTI 34
HLCrV 1 MDSRTGEIITAHQAENGVLIWTINNPLFFKTQOE 34
MYMIV 1 MDFRTGESITAAQLRSGVYTWVEVPNPLSFKIMHH 34
TYLCCNV 1 MDSRTGELLTAAQSENGVYIWTVKNPLYFKIITH 34
JayMIV 1 MDSRTGEPITAAAFNNGVFIWTVPNPLYFRVPVH 34

BC1

PYMV 150 FRAPTVKILSKQFTENDVDFSHVDYGKWERKPIR 183
CuLCrV 150 FRAPTVKLLSKQFCEKDVDFWHVGYGKWERRLVK 183
MYMV 153 FKPPITINILSKDYNENCVDFWSVEKPKPIRRLN 186
WmCSV 151 FKAPTIKILT KGYKEDCIDFWSVERGETRRRLN 184
PepYLCIV 154 FKS PKIEILSKAFNMTNIDFWTVPHSQVSRKPVQ 187
ICMV 154 FRS PKIEILSKAFTERDIDFWSVGRKAQQRKLVQ 187

B**Motivo II - AC3/BC1****AC3**

HYVMV 27 LYFKITNHAQRPFNFMN 42
TYLCCNV 27 LYFKITKHQQRPF LRN 42
TYLVGuV 27 LYFKITQHLERF LRN 42
WmCSV 27 LYFTITDHNQRPF LCQ 42
ACMV 27 LYFTITRHQQRPF LLN 42
TYLCV 27 LYFKITTAHSQRPF LFMN 42

BC1

EACMZV 184 LNPTTAHSQRPI ITHR 199
EACMCV 184 LNPTTAHSQRPI IANR 199
SACMV 184 LNPTTARSQQPI ITHR 199
MYMV 190 NQGPYTINSQRPI IMLL 205
ClGMCNV 187 LRSASARFQRPTIGP 201

C**Motivo III - AC3/BC1****AC3**

EACMCV 46 IDVQIRFNYNLRKALGMHKCFLNFRIWTRLHP 77
AsAV 46 IKIQVRFNHNLRKALGIHRCLLTFQAWTSLRP 77
PHYVV 45 FHIQVRANHNMRALGLHKA YFNFOVWTTLT 76
RhMSinV 45 YHIEVRFNHNLRKALGIHKA YFNFOVWTTSMR 76
ClGMJV 46 IQIQVRFNHNLRKELGIIHKCFLNKITYTALHP 77
EmYVV 46 IEIRIQFNYNLRKALGIHKCFLTFQI WTHLQP 77
BhYVMV 46 ITLRIHFNYNLRKALGIHKCFLTFRIWTALHP 77

BC1

RhMiMV 200 TELRPGESWASRSTVGLNQA DTESELENALHP 231
SiMBoV 200 TELRPGESWAVRSAIGINQSDAESDIENATHP 231
CdTV 200 TEIKPGESWASRSALGTGQSDADSEVENELHP 231
EuMVeV 200 IEINPGESWATRSALGMGHTNAGSDTTDALHP 231
ToYSV 200 ITIDPGETWASRSTLGIQSSSTESVENAAHP 231
MeChMV 200 IDLRPGESWASRSTLGPASITESDDAEELHP 231
ToChLDV 200 TEIRPGESWAARS AVGIGPSDADSETENEIHP 231

D**Motivo IV - AC3/BC1****AC3**

BCaMV 89 LVLLYLDRIGVVGINN 104
ToSLCV 89 LVLLYLDRLGMVICINN 104
TLCNaV 89 HVLYLDRLGMVISINN 104
TbLCZV 89 QVLYLDRLGMVISINN 104
ToLCAnV 89 HVLYLDRLGMVVSINN 104
ClGMJV 89 QVIKLYLDRLGMVISINN 104

BC1

ClLCrV 230 HPYRHLDRLGIPILDP 245
TGMV 230 HPYKHLNRLGITLIDP 245
MYMIV 243 YPLRHLHKLPEASLNP 258
EACMZV 252 YPYQRHLRLNTPELD 267
ToLCNDV 227 RPYKHLNGLDDSAIDP 242
ICMV 230 RPYSQLNRLNPDALDP 245
ClGMCNV 219 YPYRALNRLDPASLDP 234

E**Motivo V - AC3/BC1****AC3**

TYLCV 111 HVLYD**V**LENT**I**NV**T**ETHD**I**KYKFY* 134
ALCCV 111 HVLYNVLEGT**I**NVLE**C**HE**I**KFN**I**Y* 134
BCLMdV 110 YVLYD**V**LK**T**VE**C**EFT**I**Y**N**I**Q**M**K**LY* 133
PaLCrV 111 HVLYD**V**LEGT**I**NV**I**ED**H**D**I**KFN**I**Y* 134
SLCCNV 113 HVLD**N**VLENT**I**D**V**WT**T**YD**V**K**L**N**I**Y* 136
TYLCGuV 111 HVLYD**V**LEGT**I**ET**T**F**S**S**I**I**K**FN**I**Y* 134

BC1

KuMV 269 PVTRAD**I**ES**I**ET**T**INRCL**V**T**Q**RS-6-* 292
MYMIV 269 AMSKRE**I**ED**I**K**T**T**I**SKCL**I**S**Q**RS-6-* 292
EACMCV 278 TVSRKD**L**ES**I**ET**T**INRCL**I**K**T**Q**C**-6-* 301
ToLCNDV 251 S**I**TR**D**D**I**AD**I**L**K**K**T**VE**Q**C**I**K**S**NVN-7-* 274
GpMiLCV 244 Y**I**TR**E**EL**N**N**I**I**S**K**T**VD**M**CL**N**T**S**M**Q**-7-* 267
CLLCrV 262 T**L**SL**A**Q**L**N**E**L**V**K**T**T**V**Q**E**C**I**NN**C**V-8-* 285
SIGMFV 262 T**M**SL**G**Q**L**N**E**L**V**R**T**T**V**Q**E**C**I**NN**N**C**Q**-8-* 285
CoGMV 262 T**L**SR**A**Q**L**Q**E**L**V**N**S**T**V**D**L**C**I**NR**N**CT-7-* 285

Figura 19. Comparación de la proteína AC3 con la región C-terminal de BC1.

Las comparaciones entre las proteínas BC1 y AC3 fueron obtenidas usando la herramienta PSI-BLAST y se ajustaron manualmente para maximizar las correspondencias. Los residuos entre BC1 y AC3 que son idénticos (negritas) y los similares (normales) son resaltados en colores. Generalmente el color usado para resaltar es el amarillo, si en una misma posición del alineamiento existen otros aminoácidos idénticos y/o similares se usan otros colores como el morado en segunda instancia y el azul en tercera instancia. Las cajas indican bloques conservados dentro de los motivos. Cuatro o más aminoácidos idénticos y consecutivos forman un bloque, dentro de este bloque es permitido un aminoácido similar. Los cinco motivos se enumeraron consecutivamente a partir del N-terminal de la proteína AC3, y son presentado en los paneles del A al E. Los acrónimos de los nombres de los virus están coloreados de acuerdo al LP que

Al mapear los cinco motivos conservados entre la proteína AC3 y el C-terminal de BC1 encontramos que las distancias entre un motivo y otro en la proteína AC3 son similares a las distancias entre un motivo y otro de la región C-terminal de BC1 (Fig. 20a). El hecho de que los motivos estén conservados entre las dos proteínas y que los arreglos de los motivos sean similares nos indica que posiblemente la región C-terminal de BC1 está relacionada con AC3. El conjunto de datos presentados nos permitieron formular una hipótesis sobre el origen del gen BC1, nosotros creemos, como lo muestran los datos, que la región AC3 dió origen a la región C-terminal de BC1 (del aminoácido 150 al final de la proteína). Si consideramos que los aminoácidos finales de las dos proteínas son homólogos y que las proteínas BC1 tienen un tamaño entre 290-300 aa, entonces suponemos que la región que dió origen a la proteína BC1 completa es la región del componente A que se extiende desde el codón de paro del gen AC3 y 885pb (295aa) río arriba de este codón (Fig. 20B y Fig. 26).

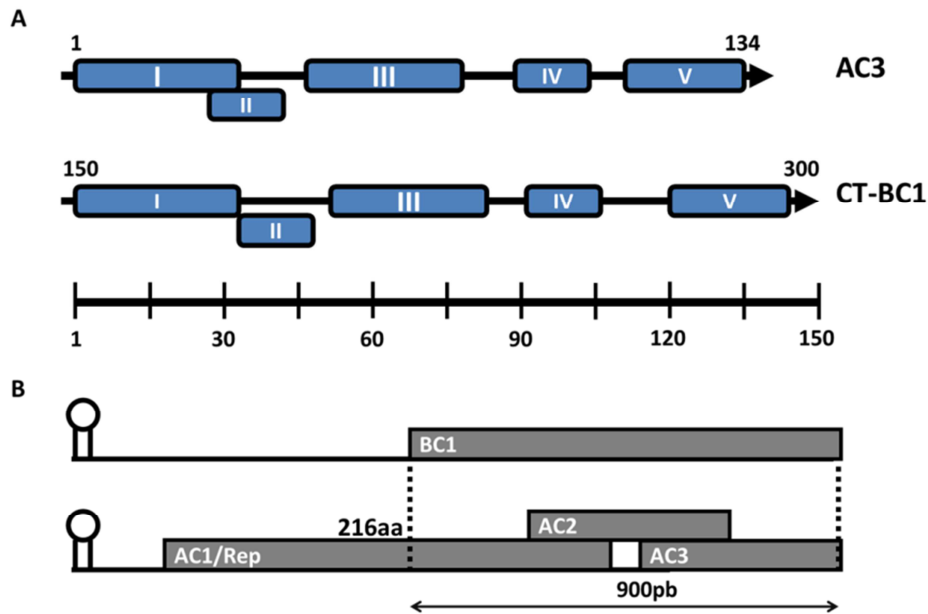


Figura 20. Mapeo de los motivos compartidos entre AC3 y BC1 e hipótesis del origen de BC1. En el panel A se muestran los motivos encontrados mediante las comparaciones con PSI-BLAST, estos son similares en secuencia y en el arreglo modular. El panel B ilustra la hipótesis sobre la región del componente A que dio origen al ORF de BC1.

Análisis comparativo entre AC2 y BC1

Suponiendo que la hipótesis sobre el origen de la proteína BC1 es cierta, nosotros decidimos realizar comparaciones entre las proteínas AC2 con la región central de BC1 y comparamos las proteínas AC1 con la región N-terminal de BC1. La figura 26 muestra qué regiones del ADN-A dieron origen a BC1.

La estrategia de comparar las proteínas AC2 en contra de la región central de la proteína BC1 nos permitió encontrar dos motivos que conservan la secuencia (Fig. 21 panel A y B) y la distancia entre ellos (Fig. 21 panel C). El primer motivo (Motivo I – AC2/BC1) tienen una longitud de 30 aminoácidos, en el alineamiento de este motivo

existen 16 posiciones que presentan residuos idénticos entre AC2 y BC1. Dentro de este motivo encontramos dos bloques (definidos con los mismos criterios que en las comparaciones de AC3 y BC1), los dos bloques están presentes en la proteína AC2 de HLCV-Pak y en BC1 del virus RhGMV (Figura 21A). El segundo motivo (Motivo II – AC2/BC1), encontrado de la comparación entre AC2 y BC1 tiene una longitud de 15 residuos de los cuales nueve posiciones del alineamiento tienen aminoácidos idénticos, este motivo no presenta bloques de cuatro aminoácidos continuos (Figura 21B). Cuando mapeamos los motivos encontrados en sus respectivas proteínas encontramos que la distancia entre ellos es conservada, en el caso de la proteína AC2 del virus EACMZV los motivos están separados por 71 aa y en la proteína BC1 del mismo virus la distancia es de 73 aa (Fig. 21C).

Al observar la posición de los bloques mapeados nos percatamos de que el motivo III de la comparación entre AC3 y BC1 (Motivo III – AC3/BC1) se encuentra sobrepuesto con el motivo II de AC2 (Motivo II – AC2/BC1) (ver Figs. 19C y 21B), por lo que decidimos comparar el motivo de AC2 con el motivo de AC3 (“Motivo II – AC2/BC1” vs Motivo III – AC3/BC1). Esta comparación nos mostró que los motivos de AC2 y AC3 (Motivo II – AC2/BC1” y Motivo III – AC3/BC1) dieron origen a la una misma región de la proteína BC1 (Fig. 22). Con los aminoácidos que constituyen esta región en AC2 y AC3 es posible reconstruir la región de la proteína BC1 que correlaciona con estos motivos de acuerdo a la hipótesis del origen de BC1 (Fig. 22A). La posición de los motivos en cada una de las proteínas es presentada en el panel B de la figura 22. Este resultado por sí solo demuestra que la región de los motivos II y III de AC2/BC1 y AC3/BC1, respectivamente, dió origen a la región 198-208 de la proteína BC1 y si a este resultado

le sumamos el hecho de que los dos motivos de AC2/BC1 se conservan en secuencia y en distancia entonces es muy probable que la región en el componente A que codifica la proteína AC2 dió origen a la región central de la proteína BC1. De hecho si representamos los cinco motivos AC3/BC1 y los dos motivos AC2/AC3 en un mismo esquema, observaríamos que las distancias entre estos siete motivos es similar en el componente A y en el componente B, lo que sugiere fuerte mente que la región central y el C-terminal de la proteína BC1 se originó de los genes sobrepuestos AC2/AC3.

Análisis comparativo entre AC1 y la región N-terminal deBC1

Hasta este punto de la comparación entre BC1 y las proteínas sobrepuestas del componente A (AC1, AC2 y AC3) hemos encontrado y concluido que los genes AC2 y AC3 dieron origen a gran parte de la proteína BC1. Para completar este análisis falta realizamos las comparaciones entre las proteínas AC1 en contra de la región N-terminal de la proteína BC1. El análisis comparativo de las proteínas BC1 en contra de las proteínas AC1 mostró que existe un motivo compartido por estas dos proteínas. El motivo está presente en las proteínas AC1 de los virus de los tres LP y fue encontrado en las proteínas BC1 de los virus africanos. Este motivo tiene una longitud de 26 aminoácidos, en donde 14 posiciones del alineamiento tienen aminoácidos idénticos. Dentro del motivo, encontramos un bloque de cinco aminoácidos contiguos que está presente en las proteínas AC1 de los virus americanos *SiMMV* y *DeLDV*, y en la proteína BC1 del virus *EACMV* (Figura 23).

A

Motivo I - AC2/BC1

AC2

RaMV	1	MQN	SS S	SP	KR	ST	Q V	P	I	K	V	Q H	K	A	K	A	I	R	R	31								
HLCV-Pak	1	M	F S	S	P	K	D	H	C	T	P V	P	I	K	V	Q H	R	E	A	K	-R	V H	R	31				
AYVSLV	1	M	R	P	S S	P	S	K	S H	S T	Q V	P	I	K	V	Q H	R	L	A	K	S	N	R	R	31			
MYMIV	1	M	R	N	S T	P	S	K	N	H	C	S P	S	I	K	V	Q H	R	V	A	E	K	K	A	I	R	R	31
TYLCMdV	1	M	R	P	S S	P	S	T	S H	C	S QV	P	I	K	V	Q H	R	I	A	K	K	R	T	I	R	31		
EACMV	1	M	P	P	S S	P	C	T	S H	C	S PV	P	I	K	V	Q H	R	T	A	K	T	R	A	L	R	31		
CoYVV	1	M-5-	T	P	S S	P	S	R	S L	S	T	A	P	S	I	K	P	R	H	R	A	K	K	V	R	R	35	
PHYVV	1	M-5-	T	P	S S	P	S	K	L S	S	P	P	A	I	K	L	R	H	R	F	A	K	R	Q	T	R	35	
RhGMHV	1	M-5-	T	P	S S	P	L	K	N L	S	S	P	P	Q	I	K	P	R	H	R	Y	A	K	K	V	R	35	

BC1

RhGMV-Sb	102	H	F S	S	F	F S	L	K	D	P	I	P	W K	L	Y	R	V	S	D	T	N	V H	Q R	131		
BCaMV	102	H	F S	S	F	F S	L	K	D	P	I	P	W K	L	Y	K	V	T	D	S	N	V H	Q R	131		
CoGMV	102	H	F S	A	S	F	F S	L	K	D	P	I	P	W K	L	Y	R	V	S	D	T	N	V H	Q R	131	
EACMZV	100	H	F S	S	T	Y	F S	L	S	E	P	S	P	W K	L	Y	R	V	E	D	S	N	V I	E	G	129
RhYMV	101	H	F S	S	F	F S	L	K	D	T	P	W E	L	V	Y	R	V	E	D	S	N	V I	D	G	130	
MYMIV	101	H	F S	S	F	F S	L	K	D	T	P	W E	L	V	Y	R	V	E	D	S	N	V I	D	G	130	
TLYCKaV	98	H	Y S	S	Y	F S	P	K	D	P	W R	V	L	Y	R	V	D	T	N	V	V	N	G	127		
JaCMV	98	H	Y S	S	Y	F S	M	K	D	V	N	P	W R	V	L	Y	R	V	D	T	N	V	N	G	127	
PepLCIV	98	H	Y S	S	Y	F S	P	K	D	P	W K	V	L	Y	R	V	D	T	N	V	N	G	127			

B

Motivo II - AC2/BC1

AC2

ChaMV	98	S	N	P V	Q P	Q P	E S	I	G	D	S	112
ACMV	98	F G	S I	Q S	Q P	E E	G	A	G	D	S	112
CPGMV	98	S T	P V	Q P	Q P	Q E	S	N	G	D	S	112
ICMV	98	E D	T V	Q P	Q P	E E	S	V	G	T	112	
StLCV	97	E S	P I	Q L	Q P	E E	S	V	G	D	111	
TbLCKoV	98	T	N	T I	Q P	Q P	Q E	S	T	G	D	112
SLCV	95	T	N	P V	Q P	Q P	E E	S	T	K	S	109
CoYVV	100	E D	P V	Q P	Q L	A E	A S	S S	114			
CLYMV	93	E D	P V	Q P	Q P	E E	S	V	G	S	107	

BC1

SiMBoV	197	F G	P I	E I	R P	G E	S W	A A	R	211	
RhMiMV	197	F G	P I	E L	R P	G E	S W	A S	R	211	
CLCαV	197	F G	P I	E I	R P	G E	S W	A T	R	211	
ACMV	198	Y	R	P I	T V	Q P	G E	T W	A	K	212
EACMZV	198	H	R	P I	T I	L P	G E	T W	A	K	212
RhYMV	220	N	R	P I	L Q	P G	E T	W A	K	234	

C

Posición de los motivos I y II AC2/BC1 en la proteína BC1 de EACMZV

Motivo I

100 **H**YF**S**S**T**Y**F**S**I**S**E**P**S**W**K**I**L**Y**R**V**E**D**S**N**V**I**E**G**V**K**F**A**S**I**K**A**K**L**R**L**S**A**K**H**S**T**D**I**R**F**K**P**P**T**I**N**I**L**S**K**G**Y**T**K**D**C**I**D**F**W**S**V**E**K**G**E**T**
R**R**R**L**L**N**P**T**A**H**S**Q**R**P**I**T**H**R**P**I**L**P**G**E**T**W**A**T**K**S**Q**I**G**L**P**S**S**S**G**P**A**R**L**E**H**F**R**S**Q**S**M**R**M 236

Motivo II

Distancia= 71 aa

Posición de los motivos I y II AC2/BC1 en la proteína AC2 de EACMZV

Motivo I

1 **M**P**P****S**S**P**S**T****S**H**C**S**L**V**P**I**K**V**Q**H**R**T**A**K**T**R**A**V**R**R**R**R**V**D**L**E**C**G**S**F**Y**L**H**I**D**C**I**N**H**G**F**S**H**R**G**T**H**C**A**S**S**Q**E**W**R**F**L**G**N**N**K**S**P**L**F**R**N**
H**Q**P**R**Q**E**A**R**E**H**E**P**R**H**H**T**P**D**T**V**Q**P**Q**E**E**G**A**G**D**S**Q**M**F**S**Q**L**Q**L**D**L**T**A**S**D**W**S**F**L**K**S**I 135

Motivo II

Distancia= 73 aa

Figura 21. Comparación de la proteína AC2 con la región central de BC1

Las comparaciones entre las proteínas BC1 y AC2 se realizaron con PSI-BLAST y se ajustaron los alineamientos manualmente. El panel A muestra el motivo I y el panel B el motivo II. Los residuos entre BC1 y AC2 idénticos (negritas) y los similares (normales) son resaltados en color amarillo. El color morado se usa para resaltar los aminoácidos idénticos o similares que se encuentran en una misma posición del alineamiento donde existen otros aminoácidos idénticos o similares. Las cajas indican bloques conservados dentro del motivo. El panel C muestra las distancias entre los motivos en cada una de las proteínas, el color amarillo indica la posición de los dos motivos. Los acrónimos son del color de cada LP que representan. Los nombres de los virus y los números de acceso de las proteínas utilizadas están listados en la tabla 3.

A

Motivo II – AC2/BC1 sobrepuesto al Motivo III - AC3/BC1

AC2

EmYVV.AC2 99 **DPIQLQPEES**I 109
 TYLCCNV.AC2 98 **GPIQLQPEES**V 108

BC1

RhMimV.BC1 198 **GPIELRPGESW** 208
 CaLCuV.BC1 198 **GPIELQPGDSW** 208

AC3

ClGMJV.AC3 44 **DIQIQVRFNH** 54
 EmYVV.AC3 44 **DEIEIRIQFNY** 54

B

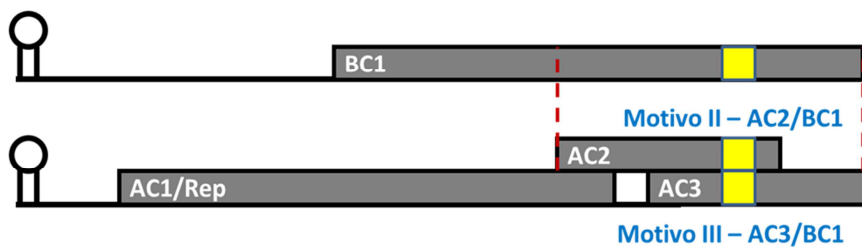


Figura 22. Alineamiento de los Motivos II – AC2/BC1 y Motivos III – AC3/BC1.

El panel A muestra el alineamiento de los motivos de II – AC2/BC1 con el motivo III – AC3/BC1. La búsqueda de los mejores ejemplos que muestran la conservación de los motivos fue realizada con BLASTP. Los aminoácido idénticos (negritas) o similares (normales) están resaltados en amarillo, en la posición 1 del alineamiento existe otro residuo idéntico (G) y se marcó en color morado, los números de acceso, los nombres y los acrónimos de los virus utilizados se enlistan en la tabla 3. El panel B ilustra la localización de los motivos (cuadros amarillos) alineados, los cuales están sobrepuestos.

Motivo I – AC1/BC1

AC1

SiMMV	239	LD	FNSK	VYS	DQV	EY	NVI	DDVAP	PHYLK	264
DeLDV	243	LD	FNNR	VYS	NEV	EY	NVI	DDVAP	QYLK	268
MYMV	244	LD	LNNK	IYS	NDAW	YN	NVI	DDV	DPHYLK	269
EACMCV	244	LD	LS	PKIYS	NDAW	YN	NVI	DDV	GPHYLK	269
TYLCV	244	LD	LS	PKVYS	NDAW	YN	NVI	DDV	DPHYLK	269
C1GMV	244	LD	LN	PKVYS	NDAW	YN	NVI	DDV	DPHYLK	269
TLCPV	250	LD	LN	PRVYS	NDAL	YN	NVI	DDV	DPHYLK	275

BC1

EACMV	07	V	T	N	N	I	N	S	K	R	T	E	Y	A	L	T	N	D	V	A	P	I	N	L	Q	32
SACMV	07	V	T	N	N	I	N	S	K	R	T	E	Y	A	L	T	N	D	A	A	P	I	N	L	Q	32
EACMCV	07	V	T	N	N	I	H	S	K	R	T	E	Y	A	L	T	N	D	A	A	P	I	H	L	Q	32
CMMaV	07	V	D	N	R	I	D	T	K	R	T	E	Y	A	L	T	N	D	A	S	P	I	Y	L	Q	32

Figura 23. Comparación de la proteína AC1 con la región N-terminal de BC1.

El análisis se realizó con PSI-BLAST y se ajustó manualmente para maximizar las coincidencias. Los residuos entre BC1 y AC1 idénticos (negritas) y los similares (normales) son resaltados en amarillo. La caja indica un bloque conservado dentro del motivo. Los acrónimos son del color de cada LP que representan. Los nombres de los virus y los números de acceso de las proteínas utilizadas están listados en la tabla 3.

Conclusión, origen de la proteína BC1

El análisis comparativo que realizamos entre las proteínas BC1 y las proteínas codificadas por los genes complementarios del ADN-A (AC1, AC2 y AC3) nos permitió concluir que los genes sobrepuestos AC1, AC2 y AC3 dieron origen a la proteína BC1. El análisis comparativo entre BC1 y AC3 mostró que con los diferentes motivos compartidos entre las proteínas AC3 diferentes virus se puede reconstruir la región C-terminal de la proteína AC3, debido a que comparten cinco motivos conservados entre ellas (Fig. 19). Estos cinco motivos no se encuentran en una proteína AC3 de un virus en particular o en una misma región C-terminal de BC1 de un virus en particular, lo que

dificultó el análisis. Sin embargo los cinco motivos encontrados conservan su arreglo modular y distancia entre cada módulo, lo que sugiere que están relacionadas (Fig. 20A). Estos datos nos permiten postular la hipótesis del origen de BC1, en donde planteamos que AC3 dió origen a la región C-terminal de BC1 (aa 150-295) y probablemente, la región que codifica al gen AC2 y la región C-terminal de AC1 dieron origen al resto de la proteína (Figs. 20B y 26). El planteamiento de esta hipótesis nos condujo a comparar la región central y N-terminal de BC1 en contra de las proteínas AC2 y el C-terminal de AC1, respectivamente. Esta estrategia nos permitió encontrar dos motivos entre AC2 y BC1 y un motivo entre AC1 y BC1. En total fueron ocho los motivos encontrados entre BC1 y las proteínas del ADN-A (AC1, AC2 y AC3). Al mapear los motivos en las proteínas BC1 y en las proteínas del componente A observamos que además de la secuencia, estos motivos conservan la distancia entre motivos aun que estos se encuentren en diferentes proteínas, como se puede observar en la figura 24. Las distancias entre los motivos mapeados en la proteína BC1 del virus EACMV y la proteína quimera del ADN-A de EACMV son similares, v.gr la distancia del Motivo I – AC1/BC1 al Motivo I – AC2/BC1 es de 67 aa en BC1 y de 61 aa en la quimera. La proteína quimera del ADN-A del virus EACMV es la unión de las proteínas AC1, AC2 y AC3 en una misma secuencia de aminoácidos. La secuencia inicia con la proteína AC1 y en la región que inicia el sobrepuesto con AC2 se substituyeron los aminoácidos de AC1 por los de AC2 y en la región en el que se sobreponen AC2 y AC3 los aminoácidos de AC2 se substituyeron por los de AC3 (Figura 24). La conservación en secuencia y la distribución de los ocho motivos apoyan nuestra hipótesis sobre el origen de BC1.

EACMV BC1

Motivi I – AC1/BC1 **Distancia 67 aa**
MDNQFTV TDNNYINSKRTEYALTNDAAAPINLQFPSSFEQATMRLKGRCKIDHIIIEYRNQVFPN

Motivi I – AC2/BC1
ATGSVIVEIRDNRVSLEDAQAAFTFPIACNVDL HYFSSTYFSISEPSFWKILYRVEDSNVIEGV

Distancia 22 aa **Motivi I – AC3/BC1** **Distancia 15 aa**
KFASIKAKLRLSLAKHSTDIR FKPPTINILSKGYTKDCIDFWSVEKGETRRRLNPTPTAHSQRP

Motivi III – AC3/BC1
ITHRP ITILPGETWATKSQIGLPSSSGPARLEHFRSQSMRMDPSTTPDLDNESTEYPYQRLHRL

HTPELDPGDSVSQTPSDSVSRKDLETLESTINKCLIKIKSEAPRPL

Proteína quimera del ADN A de EACMV

Motivi I – AC1/BC1 **Distancia 61 aa**
HNYLCGHL LDLSPKVYSNDAWYNVIDDVDPHYLKHFKEFMGAQRDWQSNTKYGKPIQIKGGIPTIF

Motivi I – AC2/BC1
LCNPGPTSSYKEFLDEEKNQSLKAWALKN MPPSSPSTSHCSLVPIKVQHRTAKTRAIRRRVDLE

Distancia 20 aa **Motivi I – AC3/BC1** **Distancia 15 aa**
CGCSFYLHIDCINH MDSRTGELITAPQARNGVFTWDITNPLYFEITDHDKRPGNMNHDIIITLQIR

Motivi III – AC3/BC1
FNHNLRKALGIHKCFLNFKIWTTLRFQTGRFLRVFRYQVLKYLDMIGVVISINTVLQAVDHVVYDV

LLNTLQVTEQHAIKFNIIY

Figura 24. Localización de los motivos presentes en BC1 y en las proteínas del ADN-A.

Cuatro de los ocho motivos encontrados se mapearon en la proteína BC1 de EACMV y en la proteína quimera de las proteínas del ADN-A. El motivo encontrado en la proteína AC1 y BC1 están resaltados en morado y el motivo encontrado entre AC2 y BC1 esta resaltado en verde, y los dos motivos resaltados en amarillo son los encontrados en las proteínas AC3 y BC1. En la proteína quimera de EACMV los aminoácidos subrayados con puntos son los residuos correspondientes a AC1 (237- 330 aa), los subrayados con una línea son los residuos de AC2 (1-50) y los subrayados con línea discontinuas son los de AC3.

Al mapear los motivos en la proteína BC1 nos dimos cuenta que en la región entre los aminoácidos 50 al 94 de la proteína BC1 no existen motivos conservados con las proteínas del componente A. Esta región se encuentra entre el Motivo I - AC1/BC1 y el Motivo I – AC2/BC1 la cual tiene una longitud de 67 aa. Ésta distancia es la de mayor número de aminoácidos que separan los diferentes motivos (Fig. 24). Es importante

resaltar que encontramos en esta región de la proteína BC1 un motivo conservado con las proteínas MP de los *Tobamovirus* (Fig. 18). Este motivo compartido entre BC1 y las MP de *Tobamovirus* es la región más conservada entre las proteínas del movimiento de la superfamilia de 30kDa y es usada para su clasificación (Mushegian & Koonin 1993). Este hallazgo nos sugiere que posiblemente la proteína BC1 se originó de forma modular, en esta situación la proteína originada por los genes AC1, AC2 y AC3 aceptó mediante recombinación el módulo característico de las proteínas del movimiento el cual le confirió a la nueva proteína la función del movimiento. Es reconocido que muchas proteínas son modulares, lo que permite la sub-funcionalización de las mismas, los módulos que tienen alguna función dentro de las proteínas son unidades discretas llamadas dominios. Los dominios son casetes de ~35 a 250 aa que típicamente retienen su estructura y propiedades bioquímicas cuando son expresados de forma aislada (Ponting & R. R. Russell 2002). Se ha mostrado que la recombinación entre dominios confiere nuevas funciones a la proteína que lo recibe (Peisajovich et al. 2010; Apic & Russell 2010), y esta recombinación de dominios sucede en patrones que se correlacionan con innovaciones evolutivas de las proteínas (Jin et al. 2009). Probablemente este escenario es aplicable al origen de la proteína BC1.

Al integrar todos los datos obtenidos de las comparaciones realizadas con BC1 y las proteínas complementarias del componente A es evidente que la región del ADN-A que codifica las proteínas AC1, AC2 y AC3 dió origen a la proteína BC1. Posiblemente esta región del ADN-A sufrió mutaciones, inserciones y/o deleciones de nucleótidos que propiciaron que el sobrepuesto de los genes se perdiera y por ende la codificación de cada una de las proteínas. Este material genético que antes codificaba para las

proteínas AC1, AC2 y AC3 posiblemente presentó cambios progresivos, como mutaciones, que propiciaron la aparición de una nueva proteína. Esta nueva proteína quizás tenía una función distinta al movimiento hasta que, por recombinación, el dominio de 30 kDa de las proteínas MP de la superfamilia de MP fue transferido a la nueva proteína.

El promotor de BC1 se originó del gen AC1

Basados en el resultado anterior, el cual indica que el gen BC1 se derivó del material genético del componente A, decidimos comparar la región del promotor BC1 contra la región del ADN-A que correlacione en posición con el promotor de BC1 en busca de zonas conservadas. Delimitamos el promotor BC1 entre el codón de inicio de la traducción y la RC de especies de los diferentes LPs y los comparamos contra la base de datos de nucleótidos del NCBI, dirigiendo la búsqueda al componente A de geminivirus, utilizando la herramienta BLASTn. El análisis arrojó que segmentos de algunos promotores BC1 están presentes en el componente A en posiciones que correlacionan con el promotor BC1. En particular, el promotor BC1 de CIGMCNV contiene una región de 51 pb, similar a un segmento del gen AC1 del componente A presente en los virus de los tres LPs (Fig. 25A). Por lo que este segmento del promotor BC1 contiene información que codifica para aminoácidos de la proteína AC1 (206-222 aa) (ver figura 25B).

Aunque son pocos los promotores BC1 que tienen regiones extendidas de similitud con el gen AC1, encontramos que en esta región está presente la CPS,

secuencia conservada en todos los promotores BC1 analizados y en todos los ORF de AC1. Como se menciono anteriormente, la secuencia CPS está presente en el marco de lectura de AC1 y codifica para los aminoácidos AARP (Figs. 8 y 25B), por lo que su localización en el componente A está conservada. Al comparar la localización de la CPS en el componente A y en el componente B encontramos que en los virus del LP Americano, excepto por algunos del clado *SLCV*, la distancia promedio en el componente B de la secuencia CPS al codón de paro de la proteína BC1 es prácticamente la misma que la distancia promedio en el componente A de la secuencia CPS al codón de paro de la proteína AC3 (Fig. 25C). Este dato sugiere que la región entre la CPS y el codón de paro de AC3 dió origen al gen BC1.

Como la región análoga del promotor BC1 en el componente A esta ocupada por el gen AC4 y parte del gen AC1, decidimos comparar los promotores BC1 contra la base de datos de aminoácidos de las proteínas AC4 y AC1 del NCBI mediante la herramienta BLASTX, la cual busca en la base de datos de proteínas usando los nucleótidos traducidos. La búsqueda mostró que en el promotor de BC1 no se encuentran codificados aminoácidos de la proteína AC4; sin embargo, encontramos que algunos promotores BC1, como el de *WiGMV*, tienen codificados aminoácidos de la proteína AC1, los cuales se encuentran en una posición análoga (Fig. 25D). La existencia del material genético del componente A en el promotor de BC1 es un fuerte indicativo de que dicho promotor se derivó evolutivamente del componente A.

El conjunto de datos evidencian una posible relación evolutiva entre el marco de lectura de la proteína AC1 y el promotor de BC1. El bloque conservado presente en el

promotor BC1 50 nucleótidos y el ORF de AC1 apoya fuertemente esta conclusión. Se podría pensar que este bloque pudo haber sido insertado en el promotor de BC1 por recombinación; sin embargo, encontrar la huella CPS nos sugiere que este escenario es poco probable ya que la huella está presente en todos los ADN-A y ADN-B analizados lo que implicaría muchos eventos de recombinación independientes. La explicación más parsimoniosa es que el promotor de BC1 se derivó evolutivamente del marco de lectura de AC1 ya que la CPS está presente en todos los componentes revisados (A y B). Sumado a esto encontramos que existen otras regiones en el promotor de BC1 que codifican para aminoácidos de la proteína AC1 (Fig.25D). Estos resultados apoyan la relación entre el promotor BC1 y la proteína AC1.

Conclusión, origen del gen BC1

Integrar los datos de las comparaciones de las proteínas complementarias del ADN-A en contra de la proteína BC1 con los datos obtenidos de las comparaciones del promotor de BC1 nos permitió formular una hipótesis sobre el origen del gen BC1. Nuestra hipótesis supone que el gen de BC1, el promotor y la región codificante, se originaron de la región complementaria del componente A. El dato más interesante que por sí sólo ilustra esta hipótesis es la conservación, en los virus americanos, de la distancia entre la CPS y el codón de paro (tomando en cuenta los 13 nts de la CPS) de las proteínas AC3 y BC1 (910 nts). El tamaño promedio de la proteína BC1 es de 295 aa, por lo tanto la región hipotética que dió origen a BC1 en el componente A se extiende 885 nts. Al mapear estos 885 nts río arriba del codón de paro de AC3 (codón homólogo al codón de paro de BC1) encontramos que el inicio de la proteína BC1

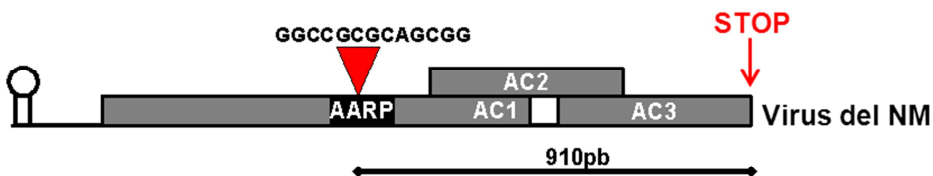
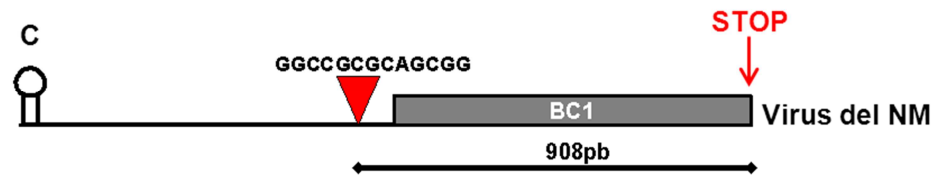
hipotética se encuentra a 9 nts de la CPS. La CPS se sitúa en el promotor de BC1 de virus americanos a 12 nts del codón de inicio de BC1, la similitud en las distancias nos sugiere fuertemente que la hipótesis planteada es verdadera. Otro dato que apoya esta hipótesis es la similitud en de los transcritos complementarios de los componentes A y B, los cuales fueron identificados por Sunter y colaboradores (1989). En este trabajo se describió que ambos componentes producen múltiples transcritos y dos de ellos presentan características compartidas. En ambos casos se produce un transcrito que inicia en la RC (nucleótido 62 en *TGMV*). No es sorprendente que los transcritos del componente A y del componente B comiencen en la región común puesto que ahí se encuentran elementos que regulan la expresión, lo que es interesante es que terminen en una posición análoga. El final del transcrito complementario del componente A se sobrepone con el final del transcrito de CP; mientras que, el final del transcrito de BC1 se sobrepone con el transcrito de BV1 en el caso del componente B. El segundo transcrito en ambos casos inicia fuera de la RC y el comienzo de estos transcritos es similar en el componente A y B; se encuentra a 135 nts río abajo del inicio del primer transcrito (Sunter & Bisaro 1989).

A

					CPS		
CIGMCNV	DNA	B	2076	ACTATCACATACTA--ACAATTCATTTATTATCAA	GGCCGCGCAGCGG	TAC	2123
CIGMCNV	DNA	A	1931	ACTAACGCCTCGATAACAATTCATTTAGGCTCAAT	GGCCGCGCAGCGG	TAC	1983
ToLCPaV	DNA	A	1941	ACTATCACCTCTATGACAATACTTTTAGGCTCAAT	GGCCGCGCAGCGG	TAT	1993
CxYVMV	DNA	A	1949	ACTATCACCTCTATCACTATACTTATGGGCTCAA	GGCCGCGCAGCGG	TAC	2001
EACMZV	DNA	A	1976	ACTATCACCTTCAAGAACAATACTATTAGGCTCAA	GGCCGCGCAGCGG	TAC	2028
ACMV	DNA	A	1941	GCTATCACCTTCTATTACAATACTATTGGGCTCAA	GGCCGCGCAGCGG	TAT	1993
HgYMV	DNA	A	1947	ACTATCACCTTCAATTACTATACTTATGGGCTCCTCC	GGCCGCGCAGCGG	TAT	1999
SiMMV	DNA	A	1878	GCTATCACCTTCAACAATTAACCTTATAGGCTCAA	GGCCGCGCAGCGG	TAC	1930
ToYVSV	DNA	A	1756	ACTATCACCTCAATTATAAATAAATAGGCTCTCC	GGCCGCGCAGCGG	TAC	1808
EuMV	DNA	A	1813	ACTATCCCTCGACTACAATACTAATAGGCTCTCT	GGCCGCGCAGCGG	TAC	1865

B

CIGMCNV	P-BC1	AAARPLINK	IVSMXDS	
CIGMCNV	AC1	237	AAARPLRPKSIVIEGVS	253
ToLCPaV	AC1	206	AAARPLRPKSIVIEGDS	222
CxYVMV	AC1	207	AAARPLRPISIVIEGDS	223
EACMZV	AC1	207	SAARPLRPNSIVLEGDS	223
ACMV	AC1	206	SAARPLRPNSIVIEGDS	222
HgYMV	AC1	207	PAARPERPISIVIEGDS	223
SiMMV	AC1	208	SAARPLRPISLIVEGDS	224
ToYVSV	AC1	209	AAARPERPISLIEGDS	225
EuMV	AC1	206	AAARPERPISIVIEGDS	222



D

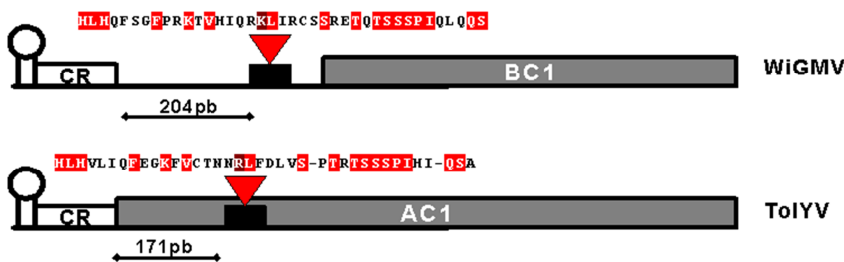


Figura 25. Análisis comparativo del promotor de BC1.

El panel A muestra el alineamiento de una región de 51 nts del promotor de BC1 de CIGMCNV (secuencia resaltada en azul), la búsqueda de regiones conservadas se realizó con BLASTn y el alineamiento se realizó manualmente. Las bases idénticas entre el promotor y el ORF se encuentran resaltadas en color rojo. La huella filogenética CPS está marcada en una caja. El panel B muestra el alineamiento del promotor BC1 traducido (resaltado en azul) con los aminoácidos de la proteína AC1 codificados en la región conservada con el promotor BC1. Los aminoácidos idénticos están resaltados en rojo y la caja señala los aminoácidos codificados por la CPS. En el panel C se ilustra la distancia que existe entre la CPS y el codón de paro (señalado por una flecha roja) de BC1 o de AC3. El panel D muestra una región del promotor de BC1 que codifica para aminoácidos de la proteína AC1 del virus *To/YV*. Los residuos idénticos están resaltados en rojo brillante y los similares en rojo oscuro. Los nombres y acrónimos de los virus utilizados están enlistados en la tabla 3.

Conclusión, el componente B de begomovirus se origina evolutivamente de un componente A

Los resultados que obtuvimos del análisis comparativo entre las proteínas y los promotores del componente B en contra del componente A nos permiten proponer que el ADN-B se originó evolutivamente de un componente A ancestral. El componente B posiblemente se originó mediante la duplicación y sub-funcionalización del gen CP que dió origen al gen BV1 y mediante la generación de *novo* del gen BC1 a partir de los genes complementarios AC1, AC2 y AC3.

La existencia de ocho bloques conservados entre las proteínas CP y BV1, y las propiedades que comparten ambas proteínas confirman la presunta homología de estas establecida previamente en el trabajo de Kikuno y colaboradores (1984). El análisis comparativo de los promotores de BV1 reveló que diversos promotores tienen una extensa similitud con promotores CP, tanto en el arreglo como en el espaciamiento de sitios de unión de factores de transcripción lo que sugiere que el promotor BV1 se

derivó de un promotor ancestral de CP. En consecuencia, el gen BV1 se derivó evolutivamente del gen CP. Esta afirmación es soportada por las similitudes que comparten los transcritos de BV1 y CP (G. Sunter & Bisaro 1989). El mecanismo más probable que pudo dar origen al gen BV1 es la duplicación del gen CP, ya que las proteínas BV1 y CP comparten bloques muy conservados y comparten la función de lanzadera.

En cuanto al gen de BC1, encontramos que la proteína contiene motivos que son conservados y se localizan en posiciones análogas en las proteínas complementarias del componente A. En la secuencias de los promotores de BC1 identificamos segmentos de 50 nts que son conservados en posición y en secuencia en ambos componentes genómicos. Dentro de esta región encontramos la secuencia palindrómica de 13 nts (CPS) la cual está presente en la mayoría de las secuencias del ADN-A y ADN-B examinadas. Integrar estos datos nos permitió concluir que el gen BC1 se originó de la región complementaria de un componente A ancestral. Esta idea es reforzada por las similitudes en los transcritos complementarios de los componentes A y B. El escenario más probable es que la proteína BC1 se originó de la región que codifica los genes sobrepuestos AC1, AC2 y AC3 y que posiblemente esta proteína nueva adquirió la función del movimiento mediante el reclutamiento del motivo característico de las proteínas del movimiento (DXR) a través de recombinación.

Métodos

Este trabajo de investigación es teórico y pretende responder aspectos importantes de la evolución de los geminivirus. La investigación se realizó de manera integrativa, conjuntando datos moleculares, mediante el análisis de genómica comparada, datos biogeográficos de los begomovirus y de su vector, datos filogenéticos y datos geológicos que nos permitieran responder tres aspectos fundamentales de la evolución de los begomovirus, ¿Cuál es el centro de origen de los BGVs?, ¿Cuál es el origen de los BGVs americanos? y ¿Cuál es el origen evolutivo del ADN-B?

Para realizar el análisis de genómica comparada de las proteínas y promotores del componente B es necesario reconstruir la filogenia de los componentes genómicos que nos guíe en la comparación de las secuencias, puesto que las comparaciones primero se realizan entre en grupo de especies relacionadas filogenéticamente y posteriormente el grupo estudiado puede compararse contra otro grupo de especies relacionado.

Las reconstrucciones filogenéticas de los componentes A y B de begomovirus, se obtuvieron alineando 63 secuencias de ADN-A y 58 de ADN-B, incluyendo el mastrevirus MSV como grupo externo en ambos casos. Las secuencias se obtuvieron de la base de datos National Center for Biotechnology Information (NCBI), los nombres, números de acceso y acrónimos utilizados se encuentran en la Tabla 2. El alineamiento múltiple se hizo con ClustalW, usando el paquete MEGA 5 (Tamura et al. 2011). Utilizamos los parámetros predefinidos por el programa (penalidad por abrir un gap: 15;

penalidad por extender el gap: 6.6; matriz de ADN: IUB). Posterior al alineamiento de las secuencias, la reconstrucción filogenética la hicimos con el método bayesiano usando el programa computacional Mr. Bayes 3.2 (Ronquist et al. 2012), con el modelo GTR+G+I. Para obtener la reconstrucción del DNA-A se corrieron un total de 2, 265, 000 árboles y el análisis lo finalizamos al obtener una desviación estándar de menor o igual a 0.01 (0.009972). Del total de los árboles se muestrearon 226,501 y de éstos se desecharon el 25% (56, 625 árboles). Por lo tanto se utilizaron 169, 876 árboles para generar la reconstrucción. Para la reconstrucción del DNA-B se corrieron un total de 1,430,000 árboles y finalizamos el análisis al obtener una desviación estándar menor o igual a 0.01 (0.010155). Del total de los arboles se muestrearon 14,300 y de estos se desecharon el 25% (3,575 árboles); por lo tanto, utilizamos 10,725 árboles para generar la reconstrucción.

Para analizar los promotores de BV1 y BC1 usamos el método denominado “Método de Análisis Filogenético Estructural” (MAFE). Este método se basa en la búsqueda de huellas filogenéticas en las secuencias de ADN de promotores homólogos de una serie filogenética. El método permite identificar elementos reguladores multipartitas que se conservan en secuencia y en la posición relativa entre los promotores. Los agrupamientos conservados de “huellas filogenéticas” (HFs) se denominan “Arreglos Modulares Conservados” (CMAs). Estos CMAs representan potenciales elementos *cis*-reguladores complejos, muy útiles además para trazar la historia evolutiva de familias génicas y genomas virales. (Argüello-Astorga & Herrera-Estrella 1998).

Para buscar las firmas de linajes primarios de las proteínas BV1 y BC1, primero identificamos los residuos conservados de cada linaje y posteriormente comparamos residuos conservados entre linajes. Las firmas se definen como los aminoácidos conservados de un LP y que no está presente en otro LP. Las comparaciones las realizamos alineando las secuencias de aminoácidos con el programa ClustalW, los parámetros usados fueron: penalidad por abrir gap 10, penalidad por extender el gap 0.20 y la matriz Blosom 62.

Para buscar homólogos de las proteínas BV1 y BC1 en la base de datos del NCBI utilizamos la herramienta "Position-Specific Iterated BLAST" (PSI-BLAST) que es un variante de BLASTP. Ésta herramienta es muy efectiva en el descubrimiento de homologías lejanas, así como en la determinación de dominios proteicos conservados. En la búsqueda contra las proteínas de la base de datos, excluimos las proteínas de los geminivirus con la intención de buscar proteínas homólogas en otros organismos. En el caso de la proteína BC1 además de buscar contra todas las proteínas de la base de datos decidimos realizar una búsqueda dirigida contra las proteínas MP de virus de plantas, las búsquedas dirigidas las realizamos usando la opción "Organism" y "Entrez Query". Para las búsquedas utilizamos las secuencias de aminoácidos de virus representativos de los tres linajes primarios y usamos los parámetros predeterminados de la herramienta (Número de palabra: 3; Matriz Blosom 62; Penalidad por abrir gap 11 y por extender el gap1).

Al no encontrar posibles homólogos de la proteína BV1 decidimos explorar la posibilidad de que BV1 y CP fuesen homólogas. Para estudiar esta hipótesis realizamos

una búsqueda con BV1 dirigida a las proteínas CP de los geminivirus utilizando la herramienta PSI-BLAST con los valores descritos arriba. Como primer paso realizamos la búsqueda con todas las proteínas BV1 de los virus utilizados en la reconstrucción filogenética. Analizando cada comparación ubicamos los bloques conservados más frecuentes, posteriormente buscamos los virus con las proteínas BV1 con el mayor número de bloques frecuentes. Con estas proteínas repetimos la búsqueda y seleccionamos las proteínas CP que presentaron mayor similitud para hacer el alineamiento manual y maximizar las coincidencias.

La búsqueda de homólogos contra toda la base de datos fue negativa por lo que decidimos realizar una búsqueda dirigida. La búsqueda dirigida para encontrar proteínas MP homólogas a BC1 se realizó utilizando la misma estrategia que con la búsqueda de BV1, en este caso un bloque fue el más frecuente. Para realizar la figura 18 buscamos las proteínas MP que presentaran la mayor similitud del único bloque conservado con las proteínas BC1 de los tres LPs. Como la longitud del bloque conservado sólo representa el 15% de la longitud de BC1, decidimos buscar si al igual que BV1 la proteína BC1 también se origina del componente A.

Usando la misma estrategia de búsqueda que utilizamos con la proteína BV1 la utilizamos con BC1. En este caso hicimos búsquedas dirigidas de la proteína BC1 en contra de cada una de las proteínas codificadas en la región complementaria del componente A. Las primeras comparaciones en donde se usó a la proteína BC1 completa mostraron que AC3 y la región C-terminal de BC1 compartían algunos bloques conservados. Con este resultado decidimos hacer búsquedas dirigidas de la

región C-terminal de BC1 en contra de las proteínas AC3. Después de diversas comparaciones logramos mapear en BC1 el aminoácido que correspondería en posición a AC3. Este residuo es la fenilalanina el cual es un aminoácido muy conservado en la posición 150 de BC1. Definimos la región C-terminal para compararla con AC3 desde el aminoácido 150 al final de BC1 (Fig. 26). Con los cinco motivos conservados entre AC3 y BC1 propusimos una hipótesis heurística que nos ayudo a dirigir las búsquedas entre AC1 vs BC1 y AC2 vs BC1. La hipótesis sugiere que la proteína BC1 se originó de la región del componente A que se extiende desde el codón de paro de AC3 hasta el aminoácido ~236 de la proteína AC1 (Fig. 26). Para hacer las comparaciones de BC1 contra AC2, utilizamos la región 100-236 de BC1 y para hacer la comparación contra AC1 utilizamos la región 1-121 de la proteína BC1 (Fig. 26).

Para hacer las comparaciones entre los promotores de BV1 y BC1 en contra la base de datos de nucleótidos del componente A de geminivirus del NCBI, nosotros utilizamos la herramienta BLASTn, usamos los parámetros predefinidos de la herramienta (Tamaño de palabra 11; puntaje por coincidencias/no coincidencias 2/-3; penalidad por abrir gap 5 y penalidad por extenderlo 2). El promotor de BV1 lo delimitamos entre el codón de inicio de la traducción de BV1 y el nucleótido 1 (sitio de corte) y el promotor de BC1 lo delimitamos entre el primer codón de BC1 y la RC.

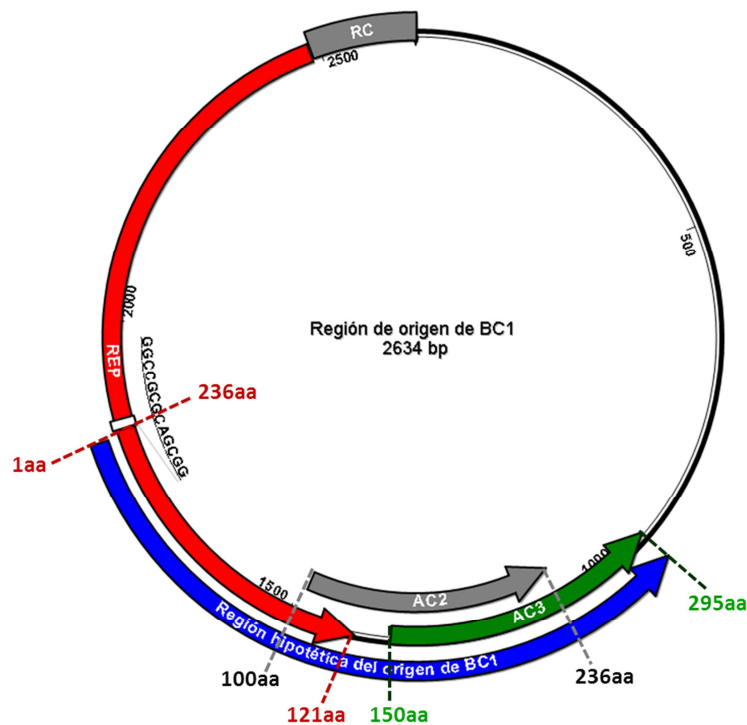


Figura 26. Región del ADN A hipotética que dió origen a BC1.

La flecha azul representa la región del componente A que posiblemente dió origen a la proteína BC1, los números en color rojo indican los aminoácidos de BC1 que se originaron de AC1, los grises son los aminoácidos de BC1 que se originaron de AC2 y los verdes de AC3.

Tabla 2 Secuencias utilizadas en la reconstrucción filogenética

Nombre del virus	Nº de Acceso ADN-A	Nº de Acceso ADN-B	Acrónimo
African cassava mosaic virus-[Nigeria]	X17095	X17096	ACMV-[NG]
Abutilon mosaic virus-[United States of America:Hawaii]	U51137	U51138	AbMV-[US:Haw]
Bean dwarf mosaic virus-[Colombia:1987]	M88179	M88180	BDMV-[CO:87]
Bean golden yellow mosaic virus-[Dominican Republic:1987]	L01635	L01636	BGMV-[DO:87]
Cabbage leaf curl virus-[United States of America:Florida:1996]	U65529	U65530	CabLCuV-[US:Flo:96]
Chino del tomate virus-Tomato [Mexico:Sinaloa H6:1983]	AF226665	N/A	CdTVTo[MX:SinH6:83]
Chino del tomate virus-Tomato [Mexico:Sinaloa B52:1983]	N/A	AF226666	CdTVTo[MX:SinB52:83]
Clerodendron golden mosaic virus-[Vietnam:Sonla:2005]	DQ641692	DQ641693	CIGMV-[VN:Son:05]
Clerodendron goleen mosaic China virus-[China:Fz7:2008]	NC_011346	NC_011347	CIGMCNV-[CN:Fz7:08]
Corchorus golden mosaic virus-[Vietnam:Hanoi:2005]	DQ641688	DQ641689	CoGMV-[VN:Han:05]

Corchorus yellow spot virus-[Mexico:Yucatan:2005]	DQ875868	DQ875869	CoYSV-[MX:Yuc:05]
Corchorus yellow vein virus-[Vietnam:Ho Binh:2000]	AY727903	AY727904	CoYVV-[VN:Ho:00]
Cotton leaf crumple virus-Arizona [Mexico:Sonora:1991]	AF480940	AF480941	CLCrVAZ[MX:Son:91]
Cowpea golden mosaic virus-[Nigeria:Nsukka:1990]	AF029217	N/A	CPGMV-[NG:Nsu:90]
Desmodium leaf distortion virus-[Mexico:Yucatan:2005]	DQ875870	DQ875871	DesLDV-[MX:Yuc:05]
East African cassava mosaic Cameroon virus-Cameroon [Cameroon:1998]	AF112354	AF112355	EACMCV-CM[CM:98]
East African cassava mosaic Kenya virus-[Kenya:Kathiana:K300:2002]	AJ717580	AJ704965	EACMKV-[KE:Kat:K300:02]
East African cassava mosaic virus-Kenya [Kenya:Boa:K48:2001]	AJ717542	AJ704949	EACMVKE[KE:Boa:K48:01]
East African cassava mosaic Zanzibar virus-[Kenya:Felunzi:K19:2001]	AJ717562	AJ704942	EACMZV-[KE:Fel:K19:01]
Euphorbia mosaic virus-[Mexico:Yucatan:2004]	DQ318937	DQ318938	EuMV-[MX:Yuc:04]
Horsegram yellow mosaic virus-[India:Coimbatore]	AJ627904	AJ627905	HgYMV-[IN:Coj]
Indian cassava mosaic virus-India [India:Maharashtra:1988]	AJ314739	AJ314740	ICMV-IN[IN:Mah:88]
Indian cassava mosaic virus	NC_001932	NC_001933	ICMV
Kudzu mosaic virus-[Vietnam:Hoabinh:2005]	DQ641690	DQ641691	KuMV-[VN:Ho:05]
Luffa yellow mosaic virus-[Vietnam]	AF509739	AF509740	LYMV-[VN]
Macroptilium mosaic Puerto Rico virus-[Puerto Rico:1990]	AY044133	AY044134	MaMPRV-[PR:90]
Mungbean yellow mosaic India virus-[India:Akola]	AY271893	AY271894	MYMIV-[IN:Ako]
Mungbean yellow mosaic virus-[India:Haryana:2001]	AY271896	N/A	MYMV-[IN:Har:01]
Mungbean yellow mosaic virus-[India:Vamban:Vigna KA34]	N/A	AJ439057	MYMV-[IN:Vam:VigKA34]
Mungbean yellow mosaic virus	NC_001983	NC_001984	MYMV
Pepper golden mosaic virus-United States of America [Mexico:Tamaulipas]	U57457	AF499442	PepGMV-US[MX:Tam]
Pepper huasteco yellow vein virus-[Mexico:Sinaloa:1988]	AY044162	AY044163	PHYVV-[MX:Sin:88]
Pepper yellow leaf curl Indonesia virus-[Indonesia:Ageratum:2005]	AB267838	AB267839	PepLCIV-A[ID:Age:05]
Potato yellow mosaic Panama virus-[Panama:Divisa:Tomato]	Y15034	Y15033	PYMPV-[PA:Div:Tom]
Rhynchosia golden mosaic Sinaloa virus-[Mexico:Sinaloa:2005]	DQ406672	DQ406673	RhGMSV-[MX:Sin:05]
Rhynchosia yellow mosaic India virus-[India:Thiruvananthapuram:2009]	NC_014902	NC_014903	RhYMIV-[IN:Thi:09]
Sida golden mosaic Costa Rica virus-[Costa Rica]	X99550	X99551	SiGMCRV-[CR]
Sida mosaic Bolivia virus-[Bolivia:2010]	HM585441	HM585442	SiMBoV-[Bo:10]
Sida micrantha mosaic virus-[Brazil:A2B2]	AJ557451	AJ557453	SiMMV-[BR:A2B2]
South African cassava mosaic virus-[South Africa]	AF155806	AF155807	SACMV-[ZA]

Soybean chlorotic blotch virus-[Nigeria:Sb19:2007]	NC_014141	NC_014142	SbCBV-[NG:Sb19:07]
Soybean mild mottle virus-[Nigeria:Sb17:2007]	NC_014140	N/A	SbMiMoV-[NG:Sb17:07]
Squash leaf curl China virus-[China:Hainan61:2005]	AM260205	AM260207	SLCCNV-[CN:Hn61:05]
Squash leaf curl Philippines virus-[Philippines:Munoz]	AB085793	AB085794	SLCPHV-PH[PH:Mun]
Squash leaf curl virus-[United States of America:Imperial Valley:1979]	M38183	M38182	SLCV-[US:Imp:79]
Sri Lankan cassava mosaic virus-India [India:Adivaram]	AJ579307	AJ579308	SLCMV-IN[IN:Adi]
Stachytarpheta leaf curl virus-[China:Hn30:2005]	AJ810156	N/A	StLCV-[CN:Hn30:05]
Jatropha curcas mosaic virus-[India:Dha:2010]	GQ924760	GQ924761	JaCMV-[IN:Dha:10]
Ramie mosaic virus-[CN:2008]	NC_010791	NC_010792	RaMV-[CN:08]
Tomato golden mosaic virus-[Brazil:Common;1984]	K02029	K02030	TGMV-[BR:Com:84]
Tomato leaf curl Gujarat virus-[India:Varanasi:2001]	AY190290	AY190291	ToLCGV-[IN:Var:01]
Tomato leaf curl New Delhi virus-[Bangladesh:Jessore: Severe:2005]	AJ875157	AJ855158	ToLCNDVIN[BG:Jes: Svr:05]
Tomato mild yellow leaf curl Aragua virus-[Venezuela:V10:2007]	NC_009490	NC_009491	ToMiYLCAV-[VE:V10:07]
Tomato mild mosaic virus-[Brazil:Pda58:2005]	EU710752	EU710753	ToMiMV-[BR:Pda58:05]
Tomato mosaic Havana virus-[Cuba:Quivican]	NC_003867	NC_003868	ToMHV-[CU:Qui]
Tomato rugose mosaic virus-[Brazil:Uberlandia 1:1996]	AF291705	AF291706	ToRMV-[BR:Ube1:96]
Tomato yellow leaf curl Kanchanaburi virus-[Thailand:Kanchanaburi 1:2001]	AF511529	AF511528	TYLCKaV-[TH:Kan1:01]
Tomato yellow leaf curl Thailand virus-A [Thailand:1]	X63015	X63016	TYLCTHV-A[TH:1]
Tomato yellow leaf curl virus-[China:JSNJ1:2009]	FN256259	N/A	TYLC-[CN:JSNJ1:09]
Tomato yellow leaf curl virus-Mild [Spain:CB1/99:2009]	AJ519441	N/A	TYLC-Mld[SPAIN:CB1/99:09]
Tomato yellow margin leaf curl virus-[Venezuela:Merida:2004]	AY508993	AY508994	TYMLCV-[VE:Mer57:04]
Tomato yellow spot virus-[Brazil:Bicas 2:1999]	DQ336350	DQ336351	ToYSV-[BR:Bic2:99]
Velvet bean severe mosaic virus-[India:Uttar:2009]	FN543425	FN543426	VeBSMV-[IN:Ut:09]
Watermelon chlorotic stunt virus-[Iran:1997]	AJ245652	AJ245653	WmCSV-[IR:97]
Wissadula golden mosaic St Thomas Virus-[Jamaica:Albion:2009]	GQ355488	NC_010951	WiGMV-[Ja:Al:09]
Maize streak virus-[Cameroon:Bambui:2009]	FM210279	N/A	MSV-[CM:Ba:09]

N/A=No aplica

Tabla 3 Secuencias de aminoácidos y nucleótidos utilizadas en las figuras presentadas

Nombre del virus	Proteína/ADN	Nº de Acceso Gen Bank	Acrónimo
Figura 10			
<i>Squash leaf curl virus</i>	BV1	ADJ38384	SLCV
<i>Chino del tomate virus</i>	BV1	AF226666	CdTV
<i>Corchorus yellow vein virus</i>	BV1	AY727904	CYVV
<i>Tomato yellow leaf curl Kanchanaburi virus</i>	BV1	AF511527	TYLCKaV
<i>Pepper yellow leaf curl Indonesia virus</i>	BV1	AB267839	PepYLCIV
<i>Tomato leaf curl New Delhi virus</i>	BV1	AD041086	ToLCNDV
<i>Mungbean yellow mosaic India virus</i>	BV1	AY271894	MYMIV
<i>Velvet bean severe mosaic virus</i>	BV1	NC_013415	VeBSMV
<i>East African cassava mosaic Cameroon virus</i>	BV1	AF112355	EACMCV
<i>South African cassava mosaic virus</i>	BV1	AF155807	SACMV
Figura 11			
<i>Desmodium leaf distortion virus</i>	BC1	YP_794173	DeLDV
<i>Bean golden yellow mosaic virus</i>	BC1	AF173556	BGYMV
<i>Corchorus golden mosaic virus</i>	BC1	ACJ73987	CoGMV
<i>Tomato yellow leaf curl Kanchanaburi virus</i>	BC1	NP_995301	TYLCKaV
<i>Pepper yellow leaf curl Indonesia virus</i>	BC1	YP_717096	PepYLCIV
<i>Tomato leaf curl New Delhi virus</i>	BC1	ADO41087	ToLCNDV
<i>Mungbean yellow mosaic India virus</i>	BC1	AFI42064	MYMIV
<i>Velvet bean severe mosaic virus</i>	BC1	YP_00325464	VeBSMV
<i>East African cassava mosaic Cameroon virus</i>	BC1	AAX39377	EACMCV
<i>South African cassava mosaic virus</i>	BC1	NP_620668	SACMV
Figura 12			
<i>Stachytarpheta leaf curl virus</i>	ADN-A	AJ810156	StLCV
<i>Ramie mosaic virus</i>	ADN-A	NC_010791	RaMV
<i>Tomato yellow leaf curl Thailand virus</i>	ADN-A	X63015	TYLCThV
<i>Tomato leaf curl Gujarat virus</i>	ADN-A	AY190290	ToLCGV
<i>Tomato yellow leaf curl Vietnam virus</i>	ADN-A	EU189150	TYLCVV
Figura 13			
<i>Tomato yellow leaf curl virus</i>	CP	CAD57979	TYLCV
<i>East African cassava mosaic Cameroon virus</i>	CP	ACF42160	EACMCV
<i>Cotton leaf curl Gezira virus</i>	CP	CBG23014	CoLCGV
<i>Tobacco leaf curl Zimbabwe virus</i>	CP	CAJ42096	TbLCZV
<i>Bean golden yellow mosaic virus</i>	CP	AAD53081	BGYMV
<i>Corchorus yellow vein virus</i>	CP	YP_115508	CoYVV
<i>Corchorus golden mosaic virus</i>	CP	YP_001333684	CoGMV
<i>Tomato yellow leaf curl Kanchanaburi virus</i>	CP	AAO15928	TYLCKaV
<i>Tomato leaf curl Gujarat virus</i>	CP	CBZ39260	ToLCGV
<i>Kudzu mosaic virus</i>	CP	ABG26011	KuMV
Figura 15			
<i>Clerodendrum golden mosaic China virus</i>	BV1	CAZ40056	CIGMCNV
	CP	YP_002268195	
<i>Tomato leaf curl New Delhi Virus</i>	CP	CAJ90652	ToLCNDV
<i>East African cassava mosaic virus</i>	CP	CAF02060	EACMV
<i>Tomato yellow leaf curl virus</i>	CP	CAD57979	TYLCV
<i>Wissadula golden mosaic virus</i>	CP	YP_001974399	WiGMV

<i>Corchorus golden mosaic virus</i>	CP	YP_001333684	CoGMV
Figura 16			
<i>Bean calico mosaic virus</i>	ADN-B	AF110190	BCaMV
<i>Tobacco yellow crinkle virus</i>	ADN-A	NC_015627	TYCrV
<i>Tomato yellow mottle virus</i>	ADN-A	AF112981	ToYMoV
<i>Pepper golden mosaic virus</i>	ADN-A	AF149227	PepGMV
Figura 17			
<i>Tomato middle mosaic virus</i>	ADN-A	EU710752	ToMiMV
<i>Sida mosaic Bolivia virus</i>	ADN-B	HM585442	SiMBoV
Figura 18			
<i>Rhynchosia yellow mosaic India virus</i>	BC1	ADT70822	RhYMIV
<i>East Africa cassava mosaic Cameroon virus</i>	BC1	AAO13748	EACMCV
<i>Clerodendrum golden mosaic China virus</i>	BC1	YP_002268201	CIGMCNV
<i>Tomato leaf curl New Delhi virus</i>	BC1	CAI44694	ToLCNDV
<i>Abutilon mosaic virus</i>	BC1	AAB18176	AbMV
<i>Corchorus golden mosaic virus</i>	BC1	YP_001333690	CoGMV
* <i>Cucumber green mottle mosaic virus</i>	MP	ADK11681	CuGrMoMV
* <i>Cucumber mottle virus</i>	MP	YP_908762	CuMoV
* <i>Kyuri green mottle mosaic virus</i>	MP	NP_619686	KyGrMoV
* <i>Maracuja mosaic virus</i>	MP	ABD94120	MaMV
* <i>Passion fruit mosaic virus</i>	MP	YP_004465360	PaFMV
* <i>Bell pepper mottle virus</i>	MP	ABC87813	BPepMoV
* <i>Tobacco mild green mosaic virus</i>	MP	ABO64419	TobMGrMV
* <i>Hibiscus latent Fort Pierce virus</i>	MP	AAT09327	HiLFPV
Figura 19			
<i>Melon chlorotic mosaic virus</i>	AC3	YP_003828905	MChMV
<i>Bean calico mosaic virus</i>	AC3	NP_612595	BCaMV
<i>Hollyhock leaf crumple virus</i>	AC3	AAD01545	HLCrV
<i>Mungbean yellow mosaic India virus</i>	AC3	AAL09038	MYMIV
	BC1	AFI42064	
<i>Tomato yellow leaf curl China virus</i>	AC3	NP_660166	TYLCCNV
<i>Jatropha yellow mosaic virus</i>	AC3	ACI15830	JaYMIV
<i>Honeysuckle yellow vein mosaic virus</i>	AC3	BAF49371	HYVMV
<i>Tomato yellow leaf curl Guangdong virus</i>	AC3	YP_764451	TYLVGuV
<i>Watermelon chlorotic stunt virus</i>	AC3	CAB65494	WmCSV
	BC1	CAB65500	
<i>African cassava mosaic virus</i>	AC3	AEO18171	ACMV
<i>Tomato yellow leaf curl virus</i>	AC3	ADQ28120	TYLCV
<i>East African cassava mosaic Cameroon virus</i>	AC3	ACF42162	EACMCV
	BC1	AAX39377	
<i>Asystasia African virus</i>	AC3	AEI91440	AsAV
<i>Pepper Huasteco yellow vein virus</i>	AC3	NP_040322	PHYVV
<i>Rhynchosia mosaic Sinaloa virus</i>	AC3	ABD65396	RhMSinV
<i>Clerodendrum golden mosaic Jiangsu virus</i>	AC3	CAZ40064	CIGMJV
<i>Emilia yellow vein virus</i>	AC3	YP_001661459	EmYVV
<i>Bhendi yellow vein mosaic virus</i>	AC3	ACN51274	BhYVMV
<i>Tomato severe leaf curl virus</i>	AC3	AEX93205	ToSLCV
<i>Tomato leaf curl Namakely virus</i>	AC3	CAI26283	TLCNaV
<i>Tobacco leaf curl Zimbabwe virus</i>	AC3	CAM91882	TbLCZV
<i>Tomato leaf curl Anjouan virus</i>	AC3	CAM91894	ToLCAnV
<i>Ageratum leaf curl Cameroon virus</i>	AC3	CCC14822	ALCCV

<i>Bean leaf curl Madagascar virus</i>	AC3	CAM91888	BCLMdV
<i>Papaya leaf crumple virus</i>	AC3	CCD22085	PaLCRV
<i>Squash leaf curl China virus</i>	AC3	ADK78875	SLCCNV
<i>Tomato yellow leaf curl Guangdong virus</i>	AC3	ACS36702	TYLCGuV
<i>Potato yellow mosaic virus</i>	BC1	AAV33838	PYMV
<i>Cucurbit leaf crumple virus</i>	BC1	AAL78210	CuLCrV
<i>Mungbean yellow mosaic virus</i>	BC1	ABI21824	MYMV
<i>Pepper yellow leaf curl Indonesia virus</i>	BC1	YP_717096	PepYLCIV
<i>India cassava mosaic virus</i>	BC1	CAJ43624	ICMV
<i>East African cassava mosaic Zanzibar</i>	BC1	CAG29056	EACMZV
<i>South African cassava mosaic virus</i>	BC1	NP_620668	SACMV
<i>Clerodendrum golden mosaic China virus</i>	BC1	YP_002268201	CIGMCNV
<i>Rhynchosia mild mosaic virus</i>	BC1	YP_004429241	RhMiMV
<i>Sida mosaic Bolivia virus</i>	BC1	YP_004207830	SiMBoV
<i>Chino del tomate virus</i>	BC1	NP_620751	CdTV
<i>Euphorbia mosaic Venezuela virus</i>	BC1	YP_001249283	EuMVeV
<i>Tomato yellow spot virus</i>	BC1	YP_459918	ToYSV
<i>Melon chlorotic mosaic virus</i>	BC1	YP_003828910	MeChMV
<i>Tomato chlorotic leaf distortion virus</i>	BC1	AEN03368	ToChLDV
<i>Cleome leaf crumple virus</i>	BC1	CBA18085	CILCrV
<i>Tomato golden mosaic virus</i>	BC1	AEI91455	TGMV
<i>Tomato leaf curl New Delhi</i>	BC1	CAX83212	ToLCNDV
<i>Kudzu mosaic virus</i>	BC1	ACL78778	KuMV
<i>Gossypium punctatum mild leaf curl virus</i>	BC1	ACQ65843	GpMiLCV
<i>Sida golden mosaic Florida virus</i>	BC1	YP_003856014	SiGMFV
<i>Corchorus golden mosaic virus</i>	BC1	YP_001333690	CoGMV
Figura 21			
<i>Ramie mosaic virus</i>	AC2	YP_001936688	RaMV
<i>Hollyhock leaf curl virus</i>	AC2	CBY88994	HLCV-Pak
<i>Ageratum yellow vein Sri Lanka virus</i>	AC2	AFA26440	AYVSLV
<i>Mungbean yellow mosaic India virus</i>	AC2	AEX31085	MYMIV
	BC1	AAX84787	
<i>Tomato yellow leaf curl Madagascar virus</i>	AC2	CAI26272	TYLCMdV
<i>East African cassava mosaic virus</i>	AC2	CAJ78151	EACMV
<i>Corchorus yellow vein virus</i>	AC2	YP_115510	CoYVV
<i>Pepper Huasteco yellow vein virus</i>	AC2	ACY95459	PHYVV
<i>Rhynchosia golden mosaic Havana virus</i>	AC2	ADN84040	RhGMHV
<i>Chayote yellow mosaic virus</i>	AC2	NP_803407	ChaMV
<i>African cassava mosaic virus</i>	AC2	CBA13497	ACMV
	BC1	CBJ57265	
<i>Cowpea golden mosaic virus</i>	AC2	AAB87608	CPGMV
<i>India cassava mosaic virus</i>	AC2	NP_047232	ICMV
<i>Stachytarpheta leaf curl virus</i>	AC2	CAH17667	StLCV
<i>Tobacco leaf curl Kochi virus</i>	AC2	NP_808776	TbLCKoV
<i>Squash leaf curl virus</i>	AC2	ABB76212	SLCV
<i>Cleome yellow mosaic virus</i>	AC2	ADK55022	CIYMV
<i>East African cassava mosaic Zanzibar virus</i>	AC2	NP_808886	EACMZV
	BC1	CAG29052	
<i>Rhynchosia golden mosaic virus</i>	BC1	ABY65911	RhGMV
<i>Bean calico mosaic virus</i>	BC1	NP_612600	BCaMV
<i>Corchorus golden mosaic virus</i>	BC1	ACK57933	CoGMV
<i>Rhynchosia yellow mosaic virus</i>	BC1	CAQ53918	RhYMV

<i>Tomato yellow leaf curl Kanchanaburi virus</i>	BC1	NP_995301	TYLCKaV
<i>Jatropha curcas mosaic virus</i>	BC1	AED89160	JaCMV
<i>Pepper yellow leaf curl Indonesia virus</i>	BC1	YP_717096	PepYLCIV
<i>Sida mosaic Bolivia virus</i>	BC1	YP_004207818	SiMBoV
<i>Rhynchosia mild mosaic virus</i>	BC1	YP_004429241	RhMiMV
<i>Cotton leaf crumple virus</i>	BC1	AAK06853	CLCrV
Figura 22			
<i>Emilia yellow vein virus</i>	AC2	YP_001661460	EmYVV
	AC3	YP_001661459	
<i>Tomato yellow leaf curl China virus</i>	AC2	AEM36732	TYLCCNV
<i>Rhynchosia mild mosaic virus</i>	BC1	YP_004429241	RhMiMV
<i>Cabbage leaf curl virus</i>	BC1	NP_624353	CaLCuV
<i>Clerodendrum golden mosaic Jiangsu virus</i>	AC3	CAZ40064	CIGMJV
Figura 23			
<i>Sida micrantha mosaic virus</i>	AC1	ADW10639	SiMMV
<i>Desmodium leaf distortion virus</i>	AC1	YP_794171	DeLDV
<i>Mungbean yellow mosaic virus</i>	AC1	ABD67446	MYMV
<i>East African cassava mosaic Cameroon virus</i>	AC1	CAI30106	EACMCV
	BC1	AAX39377	
<i>Tomato yellow leaf curl virus</i>	AC1	CAD57982	TYLCV
<i>Clerodendrum golden mosaic virus</i>	AC1	AEA10364	CIGMV
<i>Tomato leaf curl Palampur virus</i>	AC1	ACN77053	TLCPV
<i>East African cassava mosaic virus</i>	BC1	CAG29064	EACMV
<i>South African cassava mosaic virus</i>	BC1	NP_620668	SACMV
<i>Cassava mosaic Madagascar virus</i>	BC1	YP_005352920	CMMaV
Figura 24			
<i>East African cassava mosaic virus</i>	BC1	CBJ57267	EACMV
	AC1	CAJ77948	
	AC2	CBJ57255	
	AC3	CBJ57254	
Figura 25			
<i>Clerodendrum golden mosaic China virus</i>	ADN-B	FN396963	CIGMCNV
	ADN-A	FJ011668	
	AC1	ACI24162	
<i>Tomato leaf curl Pakistan virus</i>	ADN-A	FM164938	ToLCPaV
	AC1	CAQ58591	
<i>Croton yellow vein mosaic virus</i>	ADN-A	FJ593629	CrYVMV
	AC1	ACU00721	
<i>East African cassava mosaic Zanzibar virus</i>	ADN-A	AJ717564	EACMZV
	AC1	CAJ78230	
<i>African cassava mosaic virus</i>	ADN-A	X17095	ACMV
	AC1	CAA34953	
<i>Horsegram yellow mosaic virus</i>	ADN-A	AM932425	HgYMV
	AC1	CAP69650	
<i>Sida micrantha mosaic virus</i>	ADN-A	HM585433	SiMMV
	AC1	ADV15490	
<i>Tomato yellow vein streak virus</i>	ADN-A	EF417915	ToYVSV
	AC1	AAV85989	
<i>Euphorbia mosaic virus</i>	ADN-A	FN435997	EuMV
	AC1	CBA18075	
<i>Wissadula golden mosaic virus</i>	ADN-B		WiGMV

<i>Tomato infectious yellows virus</i>	AC1	AAL82833	ToIYV
----------------------------------------	-----	----------	-------

Referencias

- Alabi, O.J. et al., 2010. Two new "legumoviruses" (genus Begomovirus) naturally infecting soybean in Nigeria. *Archives of virology*, 155(5), pp.643-56.
- Altschul, S.F. et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), pp.3389-402.
- Alves-Rodrigues I, Galão RP, Meyerhans A, Díez J . 2006. *Saccharomyces cerevisiae*: a useful model host to study fundamental biology of viral replication. *Virus Research* **120**: 49-56
- Aoki, S. & Ito, M., 2000. Molecular Phylogeny of Nicotiana (Solanaceae) Based on the Nucleotide Sequence of the matK Gene. *Plant Biology*, 2(3), pp.316-324.
- Apic, G. & Russell, R.B., 2010. Domain recombination: a workhorse for evolutionary innovation. *Science signaling*, 3(139), p.pe30.
- Arlas D., Legendre M, Seltzer V, Abergel C., Claverie J-M , 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. *Proc. Natl. Acad. Sci. USA*, **108**: 17486-17491.
- Argüello-Astorga, G.R. et al., 1994. Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. *Virology*, 203(1), pp.90–100.
- Argüello-Astorga, G.R. & Herrera-Estrella, L., 1998. Evolution of Light-Regulated Plant Promoters. *Annual review of plant physiology and plant molecular biology*, 49, pp.525-555.
- Baliji, S., Sunter, J. & Sunter, G., 2007. Transcriptional analysis of complementary sense genes in Spinach curly top virus and functional role of C2 in pathogenesis. *Molecular plant-microbe interactions*, 20(2), pp.194-206.
- De Barro, P.J. et al., 2010. Bemisia tabaci: A Statement of Species Status. *Annual review of entomology*, (July 2010), pp.1-19.
- Bartoli, G. et al., 2005. Final closure of Panama and the onset of northern hemisphere glaciation. *Earth and Planetary Science Letters*, 237(1-2), pp.33-44.
- Beijerinck, M.W., 1898. Concerning a contagium vivum fluidum as cause of the spot disease of tobacco leaves. In J. J, ed. *Phytopathological classics*. St Paul: American Phytopathological Society.
- Bisaro, D.M., 2006. Silencing suppression by geminivirus proteins. *Virology*, 344(1), pp.158-68.
- Bonato, S.L. & Salzano, F.M., 1997. A single and early migration for the peopling of the Americas supported by mitochondrial DNA sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 94(5), pp.1866-1871.
- Briddon, R.W. et al., 1996. Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. *Virology*, 219(2), pp.387-94.
- Briddon, R.W. et al., 2010. Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evolutionary Biology*, 10(97), pp.1-17.

- Briddon, R.W. et al., 1990. Geminivirus coat protein gene replacement alters insect specificity. *Virology*, 177(1), pp.85-94.
- Briddon, R.W. & Stanley, J., 2006. Subviral agents associated with plant single-stranded DNA viruses. *Virology*, 344(1), pp.198-210.
- Brown, J.K. & Bird, J., 1992. Whitefly-transmitted geminiviruses and associated disorders in the Americas and the Caribbean Basin. *Plant Disease*, 76(3), pp.220-225.
- Brown, J.K. & Idris, A.M., 2005. Genetic Differentiation of Whitefly Bemisia tabaci Mitochondrial Cytochrome Oxidase I, and Phylogeographic Concordance with the Coat Protein of the Plant Virus Genus Begomovirus. *Annals of the Entomological Society of America*, 98(6), pp.827-837.
- Campbell, B.C., Steffen-Campbell, J.D. & Gill, R.J., 1994. Evolutionary origin of whiteflies inferred from 18s rDNA sequences. *Insect Molecular Biology*, 3(2), pp.73-88.
- Castillo, A.G. et al., 2003. Dual interaction of plant PCNA with geminivirus replication accessory protein (Ren) and viral replication protein (Rep). *Virology*, 312(2), pp.381-394.
- Cazzola, M. & Skoda, R.C., 2000. Translational pathophysiology: a novel molecular mechanism of human disease. *The American Society of Hematology*, 95(11), pp.3280-3288.
- Colson, P. et al., 2010. Pepper mild mottle virus, a plant virus associated with specific immune responses, Fever, abdominal pains, and pruritus in humans. *PLoS one*, 5(4), p.e10041.
- la Cour, T. et al., 2004. Analysis and prediction of leucine-rich nuclear export signals. *Protein engineering, design & selection*, 17(6), pp.527-36.
- Cui, X.-feng et al., 2005. A Begomovirus DNAbeta-Encoded Protein Binds DNA, Functions as a Suppressor of RNA Silencing, and Targets the Cell Nucleus. *Journal of virology*, 79(16), pp.10764-10775.
- Desbiez, C. et al., 1995. Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 92(12), pp.5640-5644.
- Donson, J. et al., 1984. A putative primer for second-strand DNA synthesis of maize streak virus is virion-associated. *The EMBO journal*, 3(13), pp.3069-3073.
- Dry, I.B. et al., 1997. A novel subviral agent associated with a geminivirus: the first report of a DNA satellite. *Proceedings of the National Academy of Sciences of the United States of America*, 94(13), pp.7088-7093.
- Fontenelle, M.R. et al., 2007. Functional analysis of the naturally recombinant DNA-A of the bipartite begomovirus Tomato chlorotic mottle virus. *Virus research*, 126(1-2), pp.262-267.
- Fontes, E.P.B. et al., 1994. Geminivirus replication origins have a modular organization. *The Plant cell*, 6(3), pp.405-416.
- Fontes, E.P.B. et al., 2004. The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. *Genes & development*, 18(20), pp.2545-2556.
- Frick, D.N. & Richardson, C.C., 2001. DNA Primases. *Annual review of Biochemistry*, 70, pp.39-80.

- Frischmuth, S., Frischmuth, Thomas & Jeske, H., 1991. Transcript Mapping of Abutilon mosaic virus, a Geminivirus. *Virology*, 185, pp.596-604.
- Frischmuth, T., Zimmat, G. & Jeske, H., 1990. The nucleotide sequence of abutilon mosaic virus reveals prokaryotic as well as eukaryotic features. *Virology*, 178(2), pp.461-468.
- Gafni, Y., 2002. The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. *Physiological and Molecular Plant Pathology*, 60(5), pp.231-241.
- Gardiner, W.E. et al., 1988. Genetic analysis of Tomato golden mosaic virus: the coat protein is not required for systemic spread or symptom development. *EMBO Journal*, 7(4), pp.899-904.
- Gheerbrant, E. & Rage, J.-C., 2006. Paleobiogeography of Africa: How distinct from Gondwana and Laurasia? *Palaeogeography, Palaeoclimatology, Palaeoecology*, 241(2), pp.224-246.
- Gilbertson, R.L. et al., 2003. Limitations on geminivirus genome size imposed by plasmodesmata and virus-encoded movement protein: insights into DNA trafficking. *The Plant Cell Online*, 15(11), pp.2578-2591.
- Gomez-Castañón, R.G., 2004. *Generación de vectores de expresión para el análisis funcional de secuencias del regulador virus huasteco del chile*. Instituto Potosino de Investigación Científica y Tecnológica (IPICYT).
- Goodspeed, T.H., 1954. The genus *Nicotiana*. *Journal of the American Pharmaceutical Association*, 45(3).
- Gorbalenya, A.E. & Koonin, E.V., 1993. Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology*, 3(3), pp.419-429.
- Gregorio-Jorge, J. et al., 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 journal*, 7(275), pp.1-15.
- Gruber, A.R. et al., 2008. The Vienna RNA websuite. *Nucleic acids research*, 36(Web Server issue), pp.W70-4.
- Gutierrez, C., 1999. Geminivirus DNA replication. *Cellular and Molecular Life Sciences (CMLS)*, 56(3-4), pp.313-329.
- Gutierrez, C. et al., 2004. Geminivirus DNA replication and cell cycle interactions. *Veterinary microbiology*, 98(2), pp.111-119.
- Ha, C. et al., 2006. Corchorus yellow vein virus, a New World geminivirus from the Old World. *The Journal of general virology*, 87(Pt 4), pp.997-1003.
- Ha, C. et al., 2008. Molecular characterization of begomoviruses and DNA satellites from Vietnam: additional evidence that the New World geminiviruses were present in the Old World prior to continental separation. *The Journal of general virology*, 89(Pt 1), pp.312-326.
- Hanley-Bowdoin, L. et al., 1999. Geminiviruses: Models for Plant DNA Replication, Transcription, and Cell Cycle Regulation. *Critical Reviews in Plant Sciences*, 18(1), pp.71-106.
- Hanley-Bowdoin, L., Settlege, S.B. & Robertson, D., 2004. Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. *Molecular Plant Pathology*, 5(2), pp.149-156.

- Hao, L. et al., 2003. Geminivirus AL2 and L2 Proteins Interact with and Inactivate SNF1 Kinase. *The Plant cell*, 15(April), pp.1034-1048.
- Harmon, R.S., 2005. Geological development of Panama. In R. S. Harmon, ed. *The Río Chagres, Panama*. Springer Netherlands, pp. 45-62.
- Harrison, B.D., 1985. Advances in Geminivirus Research. *Annual Review of Phytopathology*, 23(1), pp.55-82.
- Harrison, B.D., Swanson, M.M. & Fargette, D., 2002. Begomovirus coat protein: serology, variation and functions. *Physiological and Molecular Plant Pathology*, 60(5), pp.257-271.
- Hartz, M.D., Sunter, G. & Bisaro, D.M., 1999. The tomato golden mosaic virus transactivator (TrAP) is a single-stranded DNA and zinc-binding phosphoprotein with an acidic activation domain. *Virology*, 263(1), pp.1-14.
- Heyraud-Nitschke, F. et al., 1995. Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nucleic acids research*, 23(6), pp.910-916.
- Hoogstraten, R.A., Hanson, S.F. & Maxwell, D.P., 1996. Mutational analysis of the putative nicking motif in the replication-associated protein (AC1) of bean golden mosaic geminivirus. *Molecular plant-microbe interactions*, 9(7), pp.594-599.
- Hormuzdi, S.G. & Bisaro, D.M., 1993. Genetic analysis of beet curly top virus: evidence for three virion sense genes involved in movement and regulation of single- and double-stranded DNA levels. *Virology*, 193(2), pp.900-909.
- Hormuzdi, S.G. & Bisaro, D.M., 1995. Genetic analysis of beet curly top virus: examination of the roles of L2 and L3 genes in viral pathogenesis. *Virology*, 206(2), pp.1044-1054.
- Hou, M.Y. & Gilbertson, R.L., 1996. Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. *Journal of virology*, 70(8), pp.5430-5436.
- Howarth, A.J. & Goodman, R.M., 1986. Divergence and evolution of geminivirus genomes. *Journal of Molecular Evolution*, 23(4), pp.313-319.
- Ingham, D J, Pascal, E & Lazarowitz, S.G., 1995. Both bipartite geminivirus movement proteins define viral host range, but only BL1 determines viral pathogenicity. *Virology*, 207(1), pp.191-204.
- Ingham, David J, Pascal, Erica & Lazarowitz, S.G., 1995. Both bipartite geminivirus movement proteins define viral host range, but only BL1 determines viral pathogenicity. *Virology*, 207(1), pp.194-204.
- Jin, J. et al., 2009. Eukaryotic protein domains as functional units of cellular evolution. *Science signaling*, 2(98), p.ra76.
- Kannan, S. & Zacharias, M., 2007. Folding of a DNA hairpin loop structure in explicit solvent using replica-exchange molecular dynamics simulations. *Biophysical journal*, 93(9), pp.3218-3228.
- Kannan, S. & Zacharias, M., 2011. Role of the closing base pair for d(GCA) hairpin stability: free energy analysis and folding simulations. *Nucleic acids research*, 39(19), pp.8271-8280.

- Kass, G. et al., 2006. Permeabilized mammalian cells as an experimental system for nuclear import of geminiviral karyophilic proteins and of synthetic peptides derived from their nuclear localization signal regions. *The Journal of general virology*, 87(Pt 9), pp.2709-20.
- Kasteel, D.T. et al., 1996. The movement proteins of cowpea mosaic virus and cauliflower mosaic virus induce tubular structures in plant and insect cells. *The Journal of general virology*, 77, pp.2857-64.
- Kikuno, R. et al., 1984. Sequence similarity between putative gene products of geminiviral DNAs. *Nature*, 308(5959), pp.562-562.
- Kirithi, N. & Savithri, H.S., 2003. A conserved zinc finger motif in the coat protein of Tomato leaf curl Bangalore virus is responsible for binding to ssDNA. *Archives of virology*, 148(12), pp.2369-80.
- Kitajima, E.W. & Gámez, R., 1983. Electron Microscopy of Maize Rayado Fino Virus in the Internal Organs of Its Leaf hopper Vector. *Intervirology*, 19(3), pp.129-134.
- Kong, L.J. et al., 2000. A geminivirus replication protein interacts with the retinoblastoma protein through a novel domain to determine symptoms and tissue specificity of infection in plants. *The EMBO journal*, 19(13), pp.3485-3495.
- Koonin, E.V. & Iiyina, T.V., 1992. Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *Journal of General Virology*, 73, pp.2763-2766.
- Krake, L.R., Rezaian, M A & Dry, I.B., 1998. Expression of the Tomato Leaf Curl Geminivirus C4 Gene Produces Viruslike Symptoms in Transgenic Plants. *Molecular Plant-Microbe Interactions*, 11(5), pp.413-417.
- Krupovic, M., Ravantti, J.J. & Bamford, D.H., 2009. Geminiviruses: a tale of a plasmid becoming a virus. *BMC evolutionary biology*, 9(1), p.112.
- Kunik, T. et al., 1998. Nuclear import of the capsid protein of tomato yellow leaf curl virus (TYLCV) in plant and insect cells. *The Plant journal : for cell and molecular biology*, 13(3), pp.393-399.
- Lacatus, G. & Sunter, G., 2009. The Arabidopsis PEAPOD2 transcription factor interacts with geminivirus AL2 protein and the coat protein promoter. *Virology*, 392(2), pp.196-202.
- Lazarowitz, S.G. & Beachy, R.N., 1999. Viral movement proteins as probes for intracellular and intercellular trafficking in plants. *The Plant Cell Online*, 11(4), pp.535-548.
- Lefevre, P. et al., 2011. Evolutionary time-scale of the begomoviruses: evidence from integrated sequences in the Nicotiana genome. *PloS one*, 6(5), p.e19193.
- Lewis, J.D. & Lazarowitz, S.G., 2010. Arabidopsis synaptotagmin SYTA regulates endocytosis and virus movement protein cell-to-cell transport. *Proceedings of the National Academy of Sciences of the United States of America*, 107(6), pp.2491-2496.
- Liu, H. et al., 2001. Interaction of the movement and coat proteins of Maize streak virus: implications for the transport of viral DNA. *The Journal of general virology*, 82(Pt 1), pp.35-44.
- Liu, H., Boulton, M.I., et al., 1999. Maize streak virus coat protein is karyophilic and facilitates nuclear transport of viral DNA. *Molecular plant-microbe interactions : MPMI*, 12(10), pp.894-900.

- Liu, H., Boulton, M.I. & Davies, J.W., 1997. Maize streak virus coat protein binds single- and double-stranded DNA in vitro. *The Journal of general virology*, 78 (Pt 6), pp.1265-1270.
- Liu, L., Saunders, K., et al., 1999. Bean Yellow Dwarf Virus RepA, but Not Rep, Binds to Maize Retinoblastoma Protein, and the Virus Tolerates Mutations in the Consensus Binding Motif. *Virology*, 256(2), pp.270-279.
- McDougall, I., Brown, F.H. & Fleagle, J.G., 2005. Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature*, 433(February), pp.733-736.
- McGarry, R.C. et al., 2003. A novel Arabidopsis acetyltransferase interacts with the geminivirus movement protein NSP. *The Plant Cell Online*, 15(7), pp.1605-1618.
- Medeiros, R., Resende, R. & De Ávila, A.C., 2004. The plant virus Tomato Spotted Wilt Tospovirus activates the immune system of its main insect vector, *Frankliniella occidentalis*. *Journal of virology*, 78(10), pp.4976–4982.
- Morris, B. et al., 1991. Mutagenesis of the AC3 open reading frame of African cassava mosaic virus DNA reduces DNAB replication and ameliorates disease symptoms. *Journal of General Virology*, 72, pp.1205-1213.
- Murad, L. et al., 2004. The origin and evolution of geminivirus-related DNA sequences in *Nicotiana*. *Heredity*, 92(4), pp.352-358.
- Murant, A.F. & Mayo, M.A., 1982. Satellites of plant virus. *Annual review of phytopathology*, 20, pp.49-70.
- Mushegian, A.R. & Koonin, E.V., 1993. Cell-to-cell movement of plant viruses. Insights from amino acid sequence comparisons of movement proteins and from analogies with cellular transport systems. *Archives of Virology*, 133, pp.239-257.
- Nawaz-ul-Rehman, M.S. & Fauquet, C.M., 2009. Evolution of geminiviruses and their satellites. *FEBS letters*, 583(12), pp.1825-1832.
- Okamuro, J.K. & Goldberg, R.B., 1985. Tobacco single-copy DNA is highly homologous to sequences present in the genomes of its diploid progenitors. *Molecular & General Genetics*, 198(2), pp.290-298.
- Orozco, B.M. & Hanley-Bowdoin, L., 1998. Conserved sequence and structural motifs contribute to the DNA binding and cleavage activities of a geminivirus replication protein. *The Journal of biological chemistry*, 273(38), pp.24448-24456.
- Padidam, M., Beachy, R.N. & Fauquet, C.M., 1995. Classification and identification of geminiviruses using sequence comparisons. *The Journal of general virology*, 76 (Pt 2), pp.249-263.
- Palmer, K.E. & Rybicki, E.P., 1998. The Molecular Biology of Mastreviruses. K. Maramorosch, F. A. Murphy, & A. J. Shatkin, eds. *Advances in virus research*, 50, pp.183-234.
- Pant, V. et al., 2001. Molecular characterization of the Rep protein of the blackgram isolate of Indian mungbean yellow mosaic virus. *The Journal of general virology*, 82(Pt 10), pp.2559-2567.
- Paprotka, T., Metzler, V. & Jeske, H., 2010. The first DNA 1-like alpha satellites in association with New World begomoviruses in natural infections. *Virology*, 404(2), pp.148-157.

- Pascal, Erica et al., 1994. The geminivirus BR1 movement protein binds single-stranded DNA and localizes to the cell nucleus. *The Plant cell*, 6(7), pp.995-1006.
- Patil, B.L. & Fauquet, C.M., 2010. Differential interaction between cassava mosaic geminiviruses and geminivirus satellites. *The Journal of general virology*, 91(Pt 7), pp.1871-1882.
- Peisajovich, S.G. et al., 2010. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science (New York, N.Y.)*, 328(5976), pp.368-72.
- Petty, I.T.D., Coutts, R.H.A. & Buck, K.W., 1986. Geminivirus coat protein promoter sequences can function in *Escherichia coli*. *Nucleic acids research*, 14(12), p.5113.
- Ponting, C.P. & Russell, R.R., 2002. The natural history of protein domains. *Annual review of biophysics and biomolecular structure*, 31, pp.45-71.
- Pooma, W. et al., 1996. Host and Viral Factors Determine the Dispensability of Coat Protein for Bipartite Geminivirus Systemic Movement. *Virology*, 218, pp.264-268.
- Pooma, W. & Petty, I.T.D., 1996. Tomato golden mosaic virus open reading frame AL4 is genetically distinct from its C4 analogue in monopartite geminiviruses. *Journal of General Virology*, 77, pp.1947-1951.
- Poornima Priyadarshini, C.G. et al., 2011. Functional characterization of coat protein and V2 involved in cell to cell movement of Cotton leaf curl Kokhran virus-Dabawali. *PloS one*, 6(11), p.e26929.
- Qin, S., Ward, B.M. & Lazarowitz, S.G., 1998. The bipartite geminivirus coat protein aids BR1 function in viral movement by affecting the accumulation of viral single-stranded DNA. *Journal of virology*, 72(11), pp.9247-9256.
- Rigden, J.E. et al., 1993. Mutagenesis of the virion-sense open reading frames of tomato leaf curl geminivirus. *Virology*, 193(2), pp.1001-1005.
- Rigden, J.E. et al., 1994. ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. *Virology*, 204, pp.847-850.
- Rigden, J.E. et al., 1996. Plant virus DNA replication processes in *Agrobacterium*: insight into the origins of geminiviruses? *Proceedings of the National Academy of Sciences of the United States of America*, 93(19), pp.10280-10284.
- Rojas, M.R. et al., 1998. Bean Dwarf mosaic geminivirus movement proteins recognize DNA in a form- and size-specific manner. *Cell*, 95(1), pp.105-113.
- Rojas, M.R. et al., 2005. Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. *Annual review of phytopathology*, 43, pp.361-394.
- Rojas, M.R. et al., 2001. Functional analysis of proteins involved in movement of the monopartite begomovirus, Tomato yellow leaf curl virus. *Virology*, 291(1), pp.110-125.
- Romay, G. et al., 2010. Association of an atypical alphasatellite with a bipartite New World begomovirus. *Archives of virology*, 155(11), pp.1843-1847.
- Ronquist, F. et al., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic biology*, 61, pp.1-4.

- Rubinstein, G. & Czosnek, H., 1997. Long-term association of tomato yellow leaf curl virus with its whitefly vector *Bemisia tabaci*: effect on the insect transmission capacity, longevity and fecundity. *The Journal of general virology*, 78 (Pt 10, pp.2683-2689.
- Ruiz-Medrano, R. et al., 1999. Identification of a sequence element involved in AC2-mediated transactivation of the pepper huasteco virus coat protein gene. *Virology*, 253(2), pp.162-169.
- Rybicki, E.P., 1994. A phylogenetic and evolutionary justification for three genera of Geminiviridae. *Archives of Virology*, 139(1), pp.49–77.
- Saccardo, F. et al., 2011. On the alleged origin of geminiviruses from extrachromosomal DNAs of phytoplasmas. *BMC evolutionary biology*, 11(1), p.185.
- Saeed, M. et al., 2005. A Single Complementary-Sense Transcript of a Geminiviral DNA β Satellite Is Determinant of Pathogenicity. *The American Phytopathological Society*, 18(1), pp.7-14.
- Saeed, M. et al., 2007. A monopartite begomovirus-associated DNA beta satellite substitutes for the DNA B of a bipartite begomovirus to permit systemic infection. *The Journal of general virology*, 88(Pt 10), pp.2881-2889.
- Sanderfoot, A.A., Ingham, David J & Lazarowitz, S.G., 1996. A viral movement protein as a nuclear shuttle. The geminivirus BR1 movement protein contains domains essential for interaction with BL1 and nuclear localization. *Plant physiology*, 110(1), pp.23-33.
- Sanderfoot, A.A. & Lazarowitz, S.G., 1995. Cooperation in Viral Movement: The Geminivirus BL1 Movement Protein Interacts with BR1 and Redirects It from the Nucleus to the Cell Periphery. *The Plant cell*, 7(8), pp.1185-1194.
- Sanderfoot, A.A. & Lazarowitz, S.G., 1996. Getting it together in plant virus movement: cooperative interactions between bipartite geminivirus movement proteins. *Trends in cell biology*, 6(9), pp.353-358.
- Saunders, K. et al., 2000. A unique virus complex causes *Ageratum* yellow vein disease. *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp.6890-6895.
- Saunders, K., Bedford, I.D. & Stanley, J., 2001. Pathogenicity of a Natural Recombinant Associated with *Ageratum* Yellow Vein Disease: Implications for Geminivirus Evolution and Disease Aetiology. *Virology*, 282(1), pp.38-47.
- Saunders, K. & Stanley, J., 1999. A Nanovirus-like DNA Component Associated with Yellow Vein Disease of *Ageratum conyzoides*: Evidence for Interfamilial Recombination between Plant DNA Viruses. *Virology*, 264(1), pp.145-152.
- Schlee, D., 1970. Verwandtschaftsforschung an fossilen und rezenten Aleyrodina (Insectas, Hemiptera). *Stuttgarter Beitrage zur Natiirkunde*, 213, pp.1-72.
- Selth, L.A. et al., 2005. A NAC Domain Protein Interacts with Tomato leaf curl virus Replication Accessory Protein and Enhances Viral Replication. *The Plant cell*, 17, pp.311-325.
- Selth, L.A., Randles, J.W. & Rezaian, M.Ali, 2002. *Agrobacterium tumefaciens* supports DNA replication of diverse geminivirus types. *FEBS Letters*, 516, pp.179-182.

- Settlage, S.B., Miller, A.N.N.B. & Hanley-bowdoin, L., 1996. Interactions between Geminivirus Replication Proteins. *Journal of virology*, 70(10), pp.6790-6795.
- Settlage, S.B., See, R.G. & Hanley-Bowdoin, L., 2005. Geminivirus C3 protein: replication enhancement and protein interactions. *Journal of virology*, 79(15), pp.9885-9895.
- Sharma, P. & Ikegami, M., 2009. Characterization of signals that dictate nuclear/nucleolar and cytoplasmic shuttling of the capsid protein of Tomato leaf curl Java virus associated with DNA beta satellite. *Virus research*, 144(1-2), pp.145-53.
- Shivaprasad, P.V. et al., 2005. Promoters, transcripts, and regulatory proteins of Mungbean yellow mosaic geminivirus. *Journal of virology*, 79(13), pp.8149-8163.
- Singh, V.P. & Harmon, R.S., 2005. *The Río Chagres, Panama* R. S. Harmon, ed., Berlin/Heidelberg: Springer-Verlag.
- Smith, A.G., Smith, D.G. & Funnell, B., 1994. *Atlas of Mesozoic and Cenozoic Coastlines* First Edit. A. G. Smith, D. G. Smith, & B. Funnell, eds., New York, NY: Cambridge University Press.
- Stanley, J. et al., 1992. Mutational analysis of the monopartite geminivirus beet curly top virus. *Virology*, 191(1), pp.396-405.
- Stanley, J. et al., 1986. The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *The EMBO journal*, 5(8), pp.1761-1767.
- Stevens, R.D., 2006. Historical processes enhance patterns of diversity along latitudinal gradients. *Proceedings. Biological sciences / The Royal Society*, 273(1599), pp.2283-2289.
- Sunter, G. et al., 1990. Genetic analysis of tomato golden mosaic virus: ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. *Virology*, 179(1), pp.69-77.
- Sunter, G. & Bisaro, D.M., 1992. Transactivation of geminivirus AR1 and BR1 gene expression by the viral AL2 gene product occurs at the level of transcription. *The Plant cell*, 4(10), pp.1321-1331.
- Sunter, G. & Bisaro, D.M., 1989. Transcription map of the B genome component of tomato golden mosaic virus and comparison with A component transcripts. *Virology*, 173(2), pp.647-655.
- Tamura, K. et al., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, 28(10), pp.2731-2739.
- Tao, X. & Zhou, X.-ping, 2008. Pathogenicity of a naturally occurring recombinant DNA satellite associated with tomato yellow leaf curl China virus. *The Journal of general virology*, 89(Pt 1), pp.306-311.
- Thijs, G. et al., 2002. A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. *Journal of computational biology: a journal of computational molecular cell biology*, 9(2), pp.447-464.
- Thornberry, H.H., 1966. Plant pests of importance to North American agriculture. Index of plant virus diseases. *U.S. Dep. Agric.*, (307).

- Trinks, D. et al., 2005. Suppression of RNA Silencing by a Geminivirus Nuclear Protein , AC2 , Correlates with Transactivation of Host Genes. *Journal of virology*, 79(4), pp.2517-2527.
- Vanitharani, R. et al., 2004. Differential Roles of AC2 and AC4 of Cassava Geminiviruses in Mediating Synergism and Suppression of Posttranscriptional Gene Silencing. *Journal of virology*, 78(17), pp.9487-9498.
- Vaquero, C. et al., 1997. Mapping of the RNA-binding domain of the cucumber mosaic virus movement protein. *Journal of General Virology*, 78, pp.2095-2099.
- Vieira, J., Fonseca, N.A. & Vieira, C.P., 2008. An S-RNase-based gametophytic self-incompatibility system evolved only once in eudicots. *Journal of molecular evolution*, 67(2), pp.179-190.
- Voinnet, O., Pinto, Y.M. & Baulcombe, D.C., 1999. Suppression of gene silencing: A general strategy used by diverse DNA and RNA viruses of plants. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24), pp.14147-14152.
- Wallace, J.M. & Murphy, A.M., 1938. Studies on the epidemiology of curly top in Southern Idaho, with special reference to sugar beets and weed hosts of the vector. *U.S. Dept. Agric*, (624).
- Wang, H. et al., 2003. Adenosine Kinase Is Inactivated by Geminivirus AL2 and L2 Proteins The Plant Cell. *The Plant cell*, 15(December), pp.3020-3032.
- Ward, B.M. et al., 1997. The geminivirus BL1 movement protein is associated with endoplasmic reticulum-derived tubules in developing phloem cells. *Journal of virology*, 71(5), pp.3726-3733.
- Ward, B.M. & Lazarowitz, S.G., 1999. Nuclear export in plants. Use of geminivirus movement proteins for a cell-based export assay. *The Plant cell*, 11(7), pp.1267-1276.
- Wartig, L. et al., 1997. Genetic analysis of the monopartite tomato yellow leaf curl geminivirus: roles of V1, V2, and C2 ORFs in viral pathogenesis. *Virology*, 228(2), pp.132-140.
- Wezel, R.V. et al., 2003. Contribution of the Zinc Finger to Zinc and DNA Binding by a Suppressor of Posttranscriptional Gene Silencing. *Journal of virology*, 77(1), pp.696-700.
- Wezel, R.V. et al., 2002. Differential roles of geminivirus Rep and AC4 (C4) in the induction of necrosis in *Nicotiana benthamiana*. *Molecular Plant Pathology*, 3(6), pp.461-471.
- Willis, K.J. & McElwain, J.C., 2002. Flowering plant origins. In K. J. Willis & J. C. McElwain, eds. *The Evolution of Plants*. Oxford Univ Press, pp. 156-193.
- Woodburne, M.O. & Case, J.A., 1996. Dispersal, vicariance, and the Late Cretaceous to early tertiary land mammal biogeography from South America to Australia. *Journal of Mammalian Evolution*, 3(2), pp.121-161.
- Yang, X. et al., 2011. Suppression of methylation-mediated transcriptional gene silencing by β C1-SAHH protein interaction during geminivirus-beta satellite infection. *PLoS pathogens*, 7(10), p.e1002329.
- Zhang, T. et al., 2006. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS biology*, 4(1), pp.108-118.
- Zhang, W. et al., 2001. Structure of the Maize Streak Virus Geminiate Particle. *Virology*, 279, pp.471-477.

- Zhou, Y.-C. et al., 2007. The N-terminus of the Begomovirus nuclear shuttle protein (BV1) determines virulence or avirulence in *Phaseolus vulgaris*. *Molecular plant-microbe interactions*, 20(12), pp.1523-1534.
- Zou, Z., Eibl, C. & Koop, H.-U., 2003. The stem-loop region of the tobacco psbA 5'UTR is an important determinant of mRNA stability and translation efficiency. *Molecular genetics and genomics*, 269(3), pp.340-349.