

The following article appeared in *Virus*, 11(7): 644 (2019); and may be found at: <https://doi.org/10.3390/v11070644>

This is an open access article distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license <https://creativecommons.org/licenses/by/4.0/>

Article

# A Lineage of Begomoviruses Encode Rep and AC4 Proteins of Enigmatic Ancestry: Hints on the Evolution of Geminiviruses in the New World

Sandra Iliana Torres-Herrera <sup>1,5,†</sup>, Angélica Romero-Osorio <sup>1,†</sup>, Oscar Moreno-Valenzuela <sup>2</sup>, Guillermo Pastor-Palacios <sup>3</sup>, Yair Cardenas-Conejo <sup>4</sup> , Jorge H. Ramírez-Prado <sup>2</sup> , Lina Riego-Ruiz <sup>1</sup> , Yereni Minero-García <sup>2</sup>, Salvador Ambriz-Granados <sup>1</sup> and Gerardo R. Argüello-Astorga <sup>1,\*</sup>

<sup>1</sup> División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica, A.C., Camino a la Presa de San José 2055, Lomas 4ta Secc, San Luis Potosi 78216, S.L.P., México

<sup>2</sup> Centro de Investigación Científica de Yucatán, A.C., Mérida 97000, Yucatán, México

<sup>3</sup> CONACYT-CIIDZA-Instituto Potosino de Investigación Científica y Tecnológica A.C., Camino a la Presa de San José 2055, Lomas 4ta Secc, San Luis Potosi 78216, S.L.P., México

<sup>4</sup> CONACYT-Universidad de Colima, Laboratorio de Agrobiotecnología, Carretera Los Limones-Loma de Juárez (s/n), Tecnoparque CLQ Colima 28629, Colima, México

<sup>5</sup> Facultad de Ciencias Forestales, Universidad Juárez del Estado de Durango, Río Papaloapan Esquina con Blvd Durango (s/n), Col. Valle del Sur. 34120, Durango, Dgo, México

\* Correspondence: grarguel@ipicyt.edu.mx

† These authors contributed equally to this work.

Received: 20 June 2019; Accepted: 10 July 2019; Published: 13 July 2019



**Abstract:** The begomoviruses (BGVs) are plant pathogens that evolved in the Old World during the Cretaceous and arrived to the New World (NW) in the Cenozoic era. A subgroup of NW BGVs, the “*Squash leaf curl virus* (SLCV) lineage” (S-Lin), includes viruses with unique characteristics. To get clues on the evolutionary origin of this lineage, a search for divergent members was undertaken. Four novel BGVs were characterized, including one that is basal to the group. Comparative analyses led to discover a ~670 bp genome module that is nearly exclusive of this lineage, encompassing the replication origin, the *AC4* gene, and 480 bp of the *Rep* gene. A similar DNA module was found in two curtoviruses, hence suggesting that the S-Lin ancestor acquired its distinctive genomic segment by recombination with a curtovirus. This hypothesis was definitely disproved by an in-depth sequence analysis. The search for homologs of S-Lin Rep uncover the common origin of Rep proteins encoded by diverse *Geminiviridae* genera and viral “fossils” integrated at plant genomes. In contrast, no homolog of S-Lin Rep was found in public databases. Consequently, it was concluded that the SLCV clade ancestor evolved by a recombination event between a primitive NW BGV and a virus from a hitherto unknown lineage.

**Keywords:** geminivirus evolution; replication protein; curtovirus taxonomy; endogenous viral element; replication origin; rolling-circle replication

## 1. Introduction

Geminiviruses are plant viruses that possess small genomes composed by one or two circular molecules of single-stranded DNA [1] that are individually packed into geminate virions formed by two fused icosahedral hemcapsids [2,3]. The geminiviruses constitute the largest family of plant viruses, with over 450 distinct species that occur across all world regions with favourable climates for their insect vectors [4,5]. These viruses infect a multitude of wild plants in addition to a variety of

agricultural species, causing significant reduction of food stocks and heavy economic losses in many countries [4,6]. The family *Geminiviridae* is divided into distinct genera based on the insect vector species, host range, genome organization, and phylogeny [7,8]. Nine genera are currently recognized by the International Committee on Taxonomy of Viruses (ICTV): *Becurtovirus*, *Begomovirus*, *Capulavirus*, *Curtovirus*, *Eragrovirus*, *Grabovirus*, *Mastrevirus*, *Topocuvirus*, and *Turncurtovirus* [5]. The begomoviruses (BGVs) represent the ecologically more successful subgroup, encompassing 409 recognized species [9]. BGVs infect dicotyledonous plants and are transmitted by the polyphagous vector *Bemisia tabaci*, a cosmopolitan complex of morphologically indistinguishable whitefly species [10,11]. Two major phylogenetic groups of BGVs are generally recognized on the basis of geographical distribution and genome organization: the New World (NW: the Americas) and the Old World (OW: Africa, Eurasia, the Indian subcontinent, and Oceania) BGVs [12,13]. The NW BGVs (~140 species) possess a bipartite genome (DNA-A and DNA-B) with a few reported exceptions [14–16], whereas the OW BGVs include both bipartite and monopartite species, with predominance of the latter [12,17].

The genomic component A (or DNA-A) of NW BGVs displays five open reading frames (ORFs), one in the virion strand sense (*AV1* or *CP*) encoding the coat protein, and four in the complementary sense (*AC1* or *Rep*; *AC2* or *TrAP*; *AC3* or *REN*; and *AC4*) that encode proteins involved in a variety of functions including virus DNA replication, interference of plant cell cycle, temporal regulation of viral gene expression, and suppression of host antiviral responses [1,18]. The genomic component B (or DNA-B) contains only two genes: *BV1* or *NSP*, and *BC1* or *MP*, encoding proteins involved in the intracellular, intercellular, and systemic movement of the virus [19,20]. The genomic components of bipartite BGVs share a ~180 bp sequence known as “common region” (CR), which contains the plus-strand virus replication origin (*Ori*) composed by a collection of *cis*-acting sequences 5'-associated to a “stem-loop” element, where Rep cuts the plus-strand of viral DNA to initiate its replication by a rolling-circle (RCR) mechanism [21].

The NW BGVs display some features that clearly distinguish them from their Old World counterparts: (1) the absence of an *AV2* gene [13,22]; (2) a different number and arrangement of high-affinity Rep-binding sites (“iterons”) in the *Ori* region [23,24]; (3) a PWRsMaGT conserved motif in the CP N-terminal domain [22,25]; and (4) an AVR FATDK motif in the REN protein [26]. Notwithstanding those differences, diverse lines of evidence indicate that the BGVs native to the Americas evolved from OW ancestors [12,27]. In a comprehensive study that entailed both phylogenetic reconstructions and molecular clock-based methods of comparative genome analysis, Lefeuvre et al. [28] concluded that the NW BGVs evolved from Asian ancestors that probably arrived to the Americas in the Oligocene (30–35 MYA), when a land bridge between Asia and North America, and a climate favoring an unbroken belt of vegetation between those continents, presumably existed [28]. According to this scenario, the NW BGVs' ancestors would have spread from Alaska into the two Americas, engendering several secondary BGVs lineages in the course of its evolutionary radiation. One of these clades, named after the *Squash leaf curl virus*, encompasses a number of BGV species, which are scattered from Southern USA to Argentina [29–31]. Members of the S-Lin lineage (S-Lin, for simplicity) display two significant features distinguishing them from other NW BGVs: (a) the Rep-binding sites (iterons) are more numerous and display a unique arrangement in the *Ori* [23,32], and (b) the N-terminal domain of its Rep protein exhibits low aa sequence identity (<50%) with its putative homologous from other BGV lineages [24,30]. The evolutionary and functional significance of those distinctive characteristics has not been addressed up till now, notwithstanding its evident interest for a better understanding of BGVs' evolution in the Americas. In this work, we describe the molecular characterization of four novel S-Lin viruses, which were systematically compared with all BGV sequences available at public databases. This comprehensive analysis allowed us to define with very high accuracy the distinctive features of the S-Lin members, and led to the intriguing conclusion that their common ancestor evolved by a recombination event between a virus from unknown ancestry and a primal NW begomovirus.

## 2. Materials and Methods

### 2.1. Sample Collection and Begomovirus Detection

Leaf samples of weeds with symptoms of virosis were collected during the summer of 2007 in several localities of the states of Yucatan and Morelos, as a part of a broad study to assess the diversity of BGVs in Southern México. Four field samples were selected for further analysis: (1) sample YP446, from a *Capraria biflora* (*Scrophulariaceae*) plant exhibiting yellow spots and mild leaf deformation, collected in Conkal, Yucatán (21°06'56" N; 89°53'33" W); (2) sample YP122, an *Abutilon permolle* (*Malvaceae*) plant showing mosaics in leaves, which was harvested in Suytunchen, Yucatán (18°81'17" N; 98°98'33" W); (3) sample YP534, a plant of *Jacquemontia pentantha* (*Convolvulaceae*) exhibiting mosaics and mild leaf distortion, collected in Mérida, Yucatán (21°01'42" N; 89°38'1" W); and (4) a *Vigna elegans* (*Fabaceae*) plant displaying yellow mosaics and stunting (sample Mor991) was collected in Yautepéc, Morelos (18°50'17" N; 98°98'59'03" W). Total DNA was obtained from the dried samples using a modified Dellaporta protocol [33]. To increase the quantity of the field sample DNA, the extracts were subjected to rolling circle amplification (RCA) using the TempliPhi kit (GE Healthcare, Chicago, IL, USA). Polymerase chain reaction (PCR)-detection for begomoviruses was performed using Taq DNA Polymerase master mix (New England BioLabs, USA) and paired sets of lineage-specific primers, repSL2150for/cpYMAC-rev [30] and repDGR-for/cpYMAC-rev [34]. The PCR products were cloned in pGem-T Easy vector system (Promega, Madison, WI, USA). At least four clones obtained with the repSL2150for/cpYMAC-rev set of primers were sequenced for further analysis. A preliminary identification of viral sequences based on a BLAST search revealed that samples Mor991, YP122, YP446, and YP534 contained S-Lin BGVs significantly divergent in sequence from BGVs at NCBI databases. The full DNA-A sequence of S-Lin BGVs in the samples was obtained by PCR, using the pair of degenerate primers rep2370-rev/CP-EGP70-for, which produce amplicons overlapping the first PCR product along a ~600 bp long segment, thus encompassing the complete genomic component A. The DNA-B was PCR amplified using two sets of degenerate primers, BC1290for/BV1-MSKrev and BC1290-rev/BV1310-for (Table S1), which overlap in a 400 bp segment.

### 2.2. Cloning of Full-Length Viral Genomes

After the complete sequences of the DNA-A and DNA-B components of the S-Lin BGVs present in samples YP446, YP534, and Mor991 were assembled and verified, single restriction sites in each of the genomic components of those viruses were identified to design tail-to-tail (abutted) synthetic primers containing the selected endonuclease site at its 5' end. The abutted primers are listed in Supplementary Table S1. The DNA extract of each field sample was subjected to amplification by PCR using the corresponding primers and Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) in the following conditions: 2 min at 94 °C; 35 cycles at 94, 56, and 68 °C for 30, 40, and 120 s, respectively; and a final extension of 5 min at 68 °C. The PCR products encompassing the complete viral genomic component were digested with the corresponding restriction endonuclease and subcloned into pBlueScript KS+, previously digested with the same endonuclease. After ligation and transformation in *Escherichia coli*, recombinant clones were verified by both double restriction pattern and sequencing.

### 2.3. Plant Infection Assays

*Nicotiana benthamiana* plants were inoculated using a low-pressure biolistic method. The target leaves (third to fourth leaf stage) were directly shot at either 100 to 120 psi helium pressure with tungsten particles (0.7 mm, BioRad, Hercules, CA, USA) covered with DNA-A and DNA-B of each virus (5 µg). Cloned monomeric components were treated with the corresponding endonuclease to release viral DNA prior to the biolistic inoculation. JacMYuV DNA-A was digested with *Bam*HI, DNA-B with *Hind*III; CarYSYV DNA-A with *Bst*I, DNA-B with *Sal*I; and ViMYV DNA-A with *Bam*HI, DNA-B with *Kpn*I. A total of eight seedlings were inoculated with each pair of homologous components.

Mock inoculated negative controls were included for each experimental group. The inoculated plants were maintained in a growth chamber (27 °C, daily cycle of 16 h light/8 h dark) and scored for the appearance and development of disease symptoms during 3–5 weeks. All plants, both symptomless and symptomatic, were tested for the presence of viral DNA in new leaves at 14 dpi by PCR-based detection, using both DNA-A and DNA-B specific primers.

#### 2.4. Phylogenetic Analyses

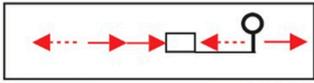
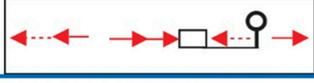
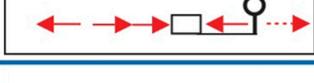
All evolutionary analyses were carried out using algorithms included in MEGA v10.0.4 [35]. The alignments of full-length DNA-A, Rep N-terminal domain (first 160 amino acid residues), and AC4 protein sequences were carried out using ClustalW algorithm with default parameters. The evolutionary history between the new viruses and related BGVs was inferred using either the neighbor-joining method [36] or the maximum likelihood method based on the Tamura–Nei model [37]. Phylogenetic analysis of Rep160 and AC4 proteins were conducted by using the maximum likelihood method based on the general reverse transcriptase model for Rep160 and the JTT matrix-based model for AC4. The substitution models were predicted by the best-fit substitution model (ML). In all cases, phylogeny tests were carried out by the bootstrap method (1000 replicates).

### 3. Results

#### 3.1. Isolation and Characterization of Novel SLCV-Lineage (S-Lin) Begomoviruses

Leaf samples of a variety of weeds exhibiting symptoms of virosis (e.g., leaf curling, yellowing, mosaics, interveinal chlorosis, growth stunting, and so on) were collected during the summer of 2007 in diverse localities of Yucatán and Morelos, México (See Section 2). The presence of BGVs was tested by PCR using a combination of both “universal” and “lineage-specific” primers designed to improve the detection of either S-Lin members or “typical” NW BGVs [30,34]. Nine out of 16 examined samples produced amplicons of the expected size (~1.4 kb) with the pair of primers repSL-2150for/cpYMACrev, specific for detection of S-Lin members. Sequencing of those PCR products showed that four of them were significantly divergent (SI < 88%) from other BGV sequences available in public databases. Those distinct viruses were isolated from *Abutilon permolle* (sample YP122), *Capraria biflora* (sample YP446), *Jacquemontia pentantha* (sample YP534), and *Vigna elegans* (sample Mor991) plants, respectively (Figure S1). The complete DNA-A was obtained by sequencing a second amplicon (~1.9 Kb) generated with the pair of primers repSL2370rev/cpEGP70 for that overlapped the first amplicon along a ~600 bp segment. The complete sequence of the cognate DNA-B of three of those BGVs was obtained by a similar procedure (see Section 2). The assembled genomic sequences were compared with those from the NCBI database using BLASTn, and pairwise nucleotide sequence identities were calculated with SDT v1.2 [17]. The analyses showed that the closest relative of each of the four isolated BGVs exhibited a full length DNA-A sequence identity lower than 90% (Figure S2), hence indicating that these viruses are separate species, according to the current ICTV taxonomic criterion for begomoviruses [17]. The new BGVs were named *Abutilon golden mosaic Yucatan virus* (AbGMYV; GenBank accession number: KC430935), *Capraria yellow spot Yucatan virus* (CarYSYV; accession no. KC426927 and KC426928), *Jacquemontia mosaic Yucatan virus* (JacMYuV; accession no. JQ821386 and JQ821387), and *Vigna yellow mosaic virus* (ViYMV; accession no. KC430936 and KC430937), respectively. The four viruses exhibited the typical genome organization of New World BGVs, with a DNA-A containing one ORF (CP) on the virion-sense strand and four ORFs (AC1–4) on the complementary-sense strand; the DNA-B included two ORFs, BV1 and BC1. In addition, all those BGVs exhibited the characteristic iteron arrangement of the SLCV clade [23,24], with two extra iterons flanking the stem-loop element at the *Ori* region (Figure 1). CarYSYV harbored six iterons with a TGGAGTCC consensus, whereas the five iterons of AbGMYV and JacMYuV displayed a TGGTGTCC sequence. In contrast, ViYMV exhibited five iterated elements with a TGGAGACC consensus, differing in the sixth nucleotide from the former BGVs iterons.

In fact, ViYMV iterons are unique in sequence among the S-Lin members. The relevant information on the novel viruses is summarized in Figure 1.

Virus name	Host	Genome components	Common Region	Closest Relative	Iteron arrangement
Abutilon golden mosaic Yucatan virus <b>AbGMYV</b>	<i>Abutilon permolle</i> ( <i>Malvaceae</i> )	<b>DNA-A</b> (2629 nt)	ND	<b>DeLDV</b> DQ875870 <b>SI= 85.6%</b>	
Capraria yellow spot Yucatan virus <b>CarYSYV</b>	<i>Capraria biflora</i> ( <i>Scrophulariaceae</i> )	<b>DNA-A</b> (2624 nt) <b>DNA-B</b> (2584 nt)	<b>154 nt</b> (SI=72%)	<b>CbLCuV</b> U65529 <b>SI=83.7%</b>	
Jacquemontia mosaic Yucatan virus <b>JacMYuV</b>	<i>Jacquemontia pentantha</i> ( <i>Convolvulaceae</i> )	<b>DNA-A</b> (2604 nt) <b>DNA-B</b> (2591 nt)	<b>190 nt</b> (SI=72%)	<b>JacYVV</b> KY617094 <b>SI=79.7%</b>	
Vigna yellow mosaic virus <b>ViYMV</b>	<i>Vigna elegans</i> ( <i>Fabaceae</i> )	<b>DNA-A</b> (2602 nt) <b>DNA-B</b> (2564 nt)	<b>192 nt</b> (SI=72%)	<b>BCaMV</b> AF11089 <b>SI=87.2%</b>	

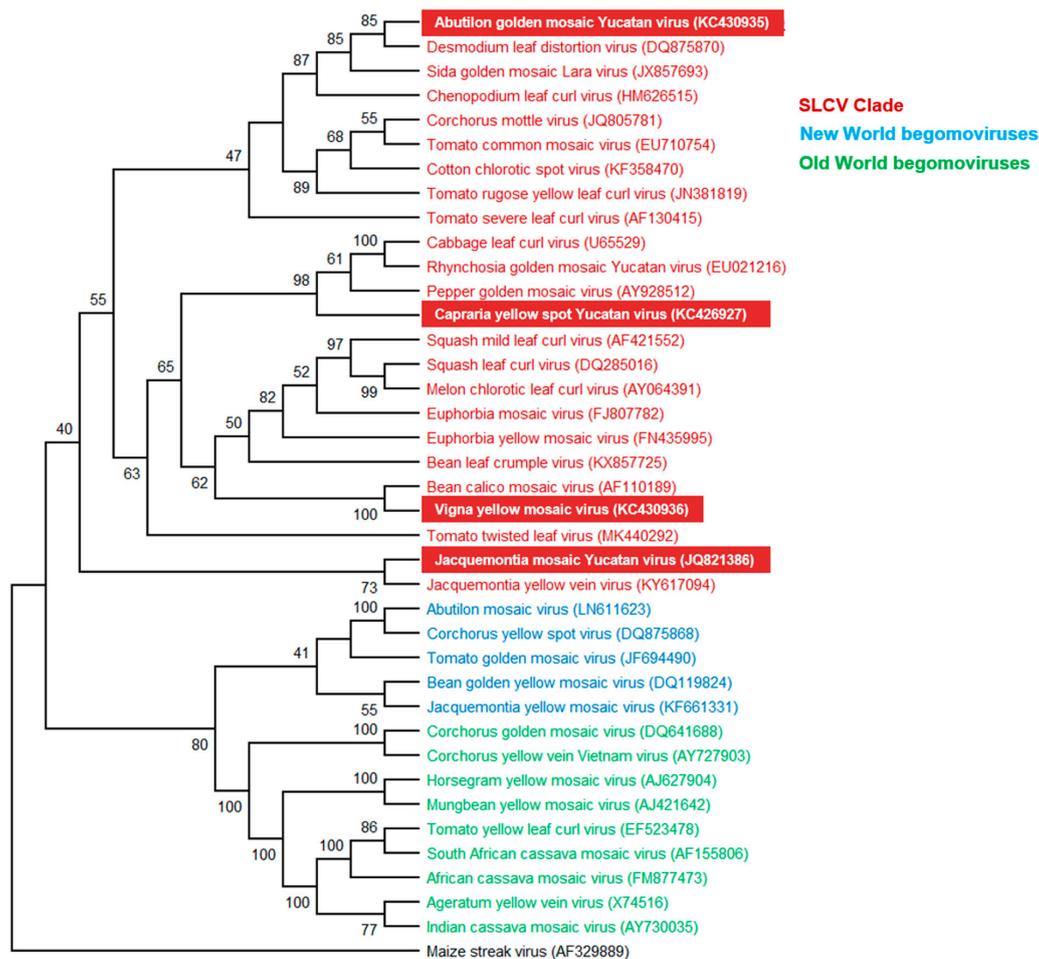
**Figure 1.** Summary of significant data on the begomoviruses described in this work. SI, sequence identity; ND, not determined (DNA-B is not known). Red arrows represent iterons (Rep-binding sites) and discontinuous arrows represent imperfect repeats; the TATA-box is represented by a square, and the conserved stem-loop element by a balloon-like figure.

### 3.2. Recombination Analysis

The DNA-A of the new BGVs was analyzed with the suite of programs for detection of recombinant breakpoints in the RDP3 package [38]. The analysis did not reveal significant evidence of recombination in CarYSYV, JacMYuV, and ViYMV. In contrast, one recombination event with breakpoints at positions 1969 and 2595 was identified in AbGMYV. One recombinant 626 nt long segment was presumably derived from *Desmodium leaf deformation virus* (DeLDV; the “minor progenitor”) encompassing the complete AC4 ORF, the 5′ third of the Rep gene, and most of the common region (Figure S3). The remaining part of AbGMYV DNA-A was apparently derived from *Corchorus yellow spot virus* (CoYSV). It is noteworthy that both putative progenitors of AbGMYV are native to Yucatan Peninsula [39,40], and that CoYSV is not a member of the S-Lin. The recombination between CoYSV and DeLDV was well supported by six methods in the RDP3 package (GENECONV,  $p = 2.649 \times 10^{-34}$ ; RDP,  $p = 3.606 \times 10^{-31}$ ; BootScan,  $P = 6.429 \times 10^{-27}$ ; MaxChi,  $p = 1.172 \times 10^{-27}$ ; Chimaera,  $p = 1.636 \times 10^{-25}$ ; 3Seq,  $p = 7.617 \times 10^{-50}$ ).

### 3.3. Phylogenetic Relationships of the New Begomoviruses

A phylogenetic tree based on the DNA-A sequences of ~90 NW BGVs and 30 OW BGVs was generated using the neighbor-joining method with 1000 bootstraps replications. The resulting dendrogram [41] confirmed that the novel BGVs belong to the SLCV lineage. In Figure 2, a simplified phylogenetic tree including 24 S-Lin viruses and 14 “typical” (i.e., non S-Lin) NW and OW BGVs is illustrated.



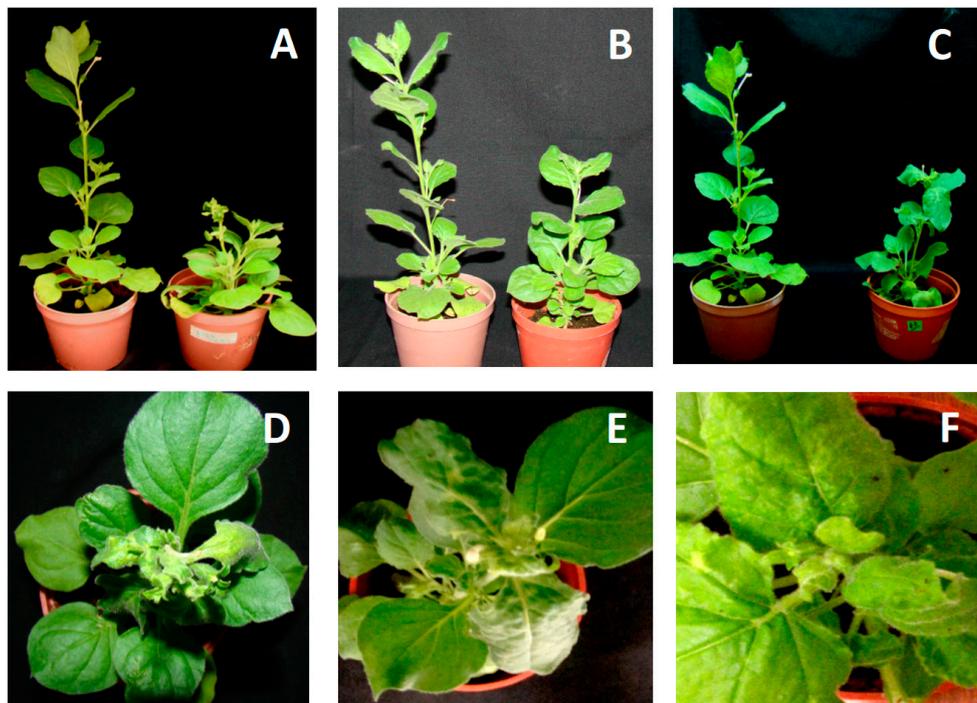
**Figure 2.** Phylogenetic relationships between the novel begomoviruses (highlighted in red background) and selected begomoviruses. Phylogenetic analysis was based on the alignment of the full-length DNA-A sequences of 24 members of the SLCV clade (highlighted in red), five begomoviruses (BGVs) from New World (highlighted in blue), nine BGVs from Old Word (highlighted in green), and one mastrevirus as outgroup (black). The phylogenetic tree was inferred using the neighbor-joining method. Bootstrap values (1000 iterations) are indicated for each node. GenBank accession numbers are indicated in parentheses.

CarYSYV grouped into a small cluster of Mesoamerican viruses that includes *Cabbage leaf curl virus* (CaLCuV) and *Pepper golden mosaic virus* (PepGMV); ViYMV was closely related to other legume-infecting BGVs, that is, *Bean calico mosaic virus* (BCaMV) and *Bean leaf crumple virus* (BLCrV); AbGMYV grouped into a subclade including both Mesoamerican and South American BGVs, while JacMYuV and *Jacquemontia yellow vein virus* (JacYVV) were basal to the clade. A phylogeny of the same DNA-A sequences that was reconstructed using the maximum likelihood method is illustrated in Figure S4. On the other hand, a phylogenetic tree based on comparisons of full-length DNA-B sequences was not utterly congruent with that reconstituted from the DNA-A. For instance, the DNA-B of JacMYuV clustered into the large AbMV clade of NW BGVs, whereas CarYSYV was closer to the PHYVV cluster (Figure S5). Incongruent phylogenies of DNA-A and DNA-B components of a bipartite begomovirus are generally explained by recombination and/or re-assortment events along the evolutionary history of that particular species [12].

### 3.4. Infectivity of Cloned DNA-A and DNA-B Components

Full-length clones of both genomic components of JacMYuV, CarYSYV, and ViYMV were generated to verify their capability to produce infections in plants. The infectious clones were obtained via PCR

amplification using back-to-back primers designed on unique restriction sites (see Section 2). Biolistic inoculation of *Nicotiana benthamiana* plants with clones of JacMYuV DNA-A and DNA-B produced clear symptoms at 10 dpi in all tested plants (8/8). Disease signs included strong shortening of internodes, and yellowish and severe deformation of new leaves (Figure 3A). Symptoms did not diminish with time, and plants did not develop flowers even after 40 dpi. In contrast, six out of eight plants bombarded with CarYSYV clones developed mild symptoms, including a slight delay in growth and mild leaf curling (Figure 3B), which lessened gradually after 20 dpi, suggesting remission. Finally, the clones of ViMYV produced symptoms in six out of eight inoculated plants, which exhibited mild chlorosis and deformation of new leaves at 12 dpi, although remission of symptoms was noticed after 25 dpi (Figure 3C). These results showed that the three tested BGVs have different capabilities to induce symptoms in the common host *N. benthamiana*, and confirmed the complementary nature of the isolated viral genomic components. This was an important issue in the case of CarYSYV, whose common region (CR) exhibited a low conservation in sequence (SI = 72%) between cognate genomic components, as previously indicated (see Figure 1).

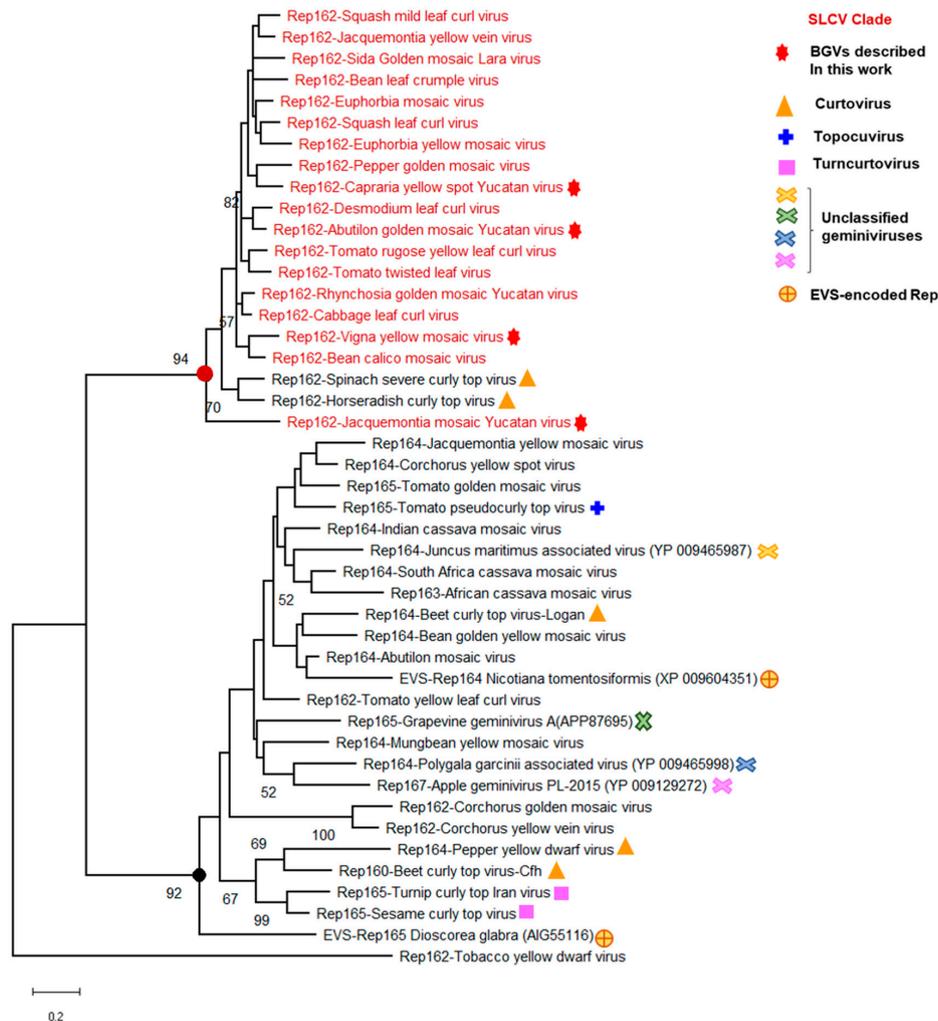


**Figure 3.** *Nicotiana benthamiana* plants inoculated with DNA-A and DNA-B of (A,D) JacMYuV, (B,E) CarYMYV, and (C,F) ViMYV. The symptomatic plants (28 dpi) (right side of photography) are compared with mock-inoculated plants (left side).

### 3.5. Identification of a Distinctive Rep Domain Encoded by SLCV-Lin BGVs

To define with higher accuracy the genomic and molecular hallmarks of the SLCV lineage members, the intergenic region and the predicted proteins encoded in the DNA-A of all its members were systematically compared with those from other NW BGVs. This broad analysis revealed a number of S-Lin specific motifs in the Rep N-terminal domain (aa 1–160), but not in other domains of that protein. The expansion of the comparative analysis to include Rep proteins encoded by OW BGVs from distinct lineages, such as the so-called Corchoviruses [22,25], Legumoviruses [42], Sweepoviruses [12,17], and other BGVs indigenous to OW landmasses, confirmed that the identified S-Lin Rep motifs are actually unique among the replication proteins encoded by BGVs [30]. Consequently, the comparative analysis was extended to embrace the Rep proteins of all *Geminiviridae* genera, looking for a plausible ancestor of the S-Lin Rep. A cladogram derived from the comparisons of Rep N-domain (aa 1–160) encoded by

representative members from different geminivirus genera is illustrated in Figure 4. Several data of this phylogeny are noteworthy: (1) the Rep160 of SLCV-Lin viruses form a cluster clearly separated from the main branch of the equivalent Rep domain of geminiviruses belonging to four genera: *Begomovirus*, *Curtovirus*, *Topocuvirus*, and *Turncurtovirus*; (2) The Rep160 of JacMYuV is basal to the SLCV cluster; (3) the Rep160 of two curtoviruses, *Horseradish curly top virus* (HrCTV) [43] and *Spinach severe curly top virus* (SpSCTV) [44], are clustered within the S-Lin Rep branch. This observation naturally suggests that the distinctive Rep domain of S-Lin BGVs could be derived from a curtovirus related to HrCTV, a presumption that was later examined at close detail (see below).

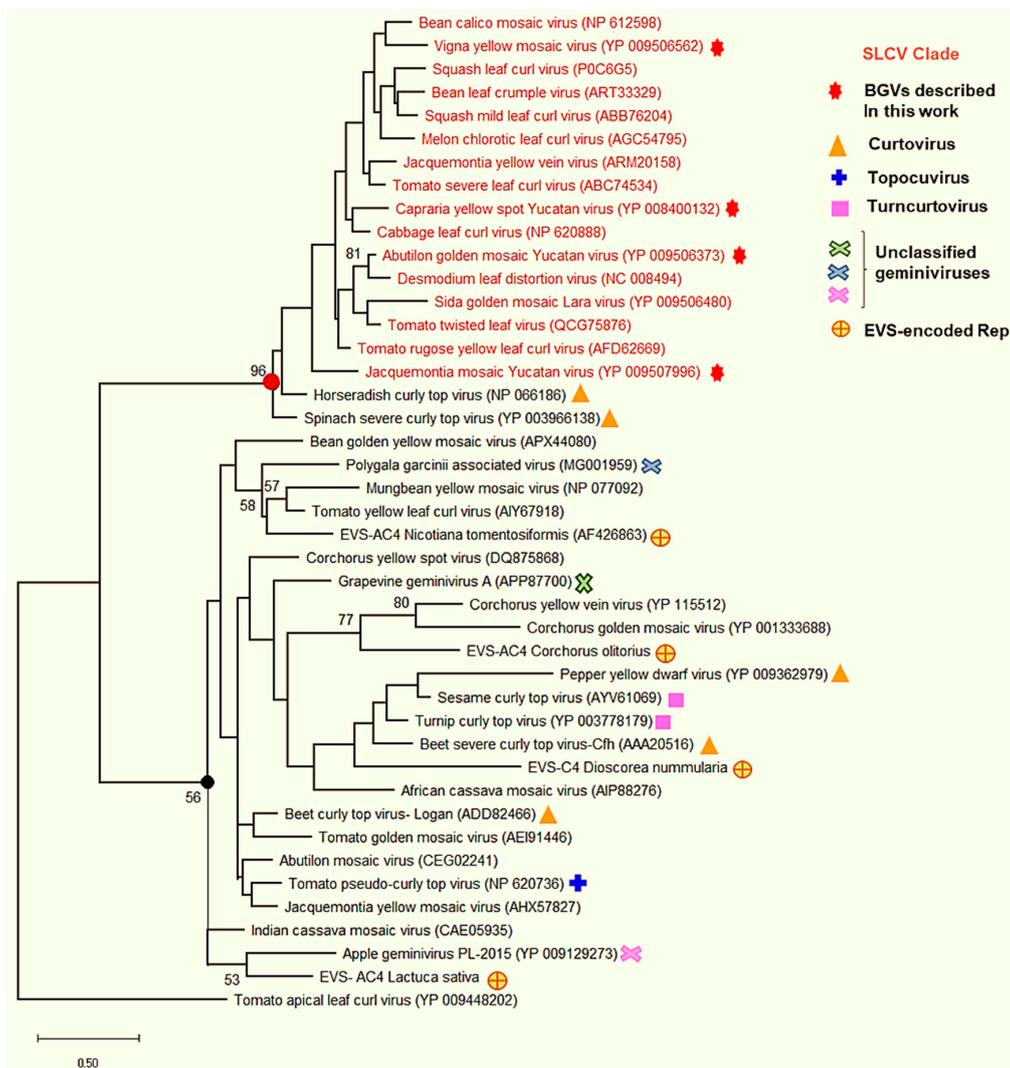


**Figure 4.** Phylogenetic relationships between the Rep N-terminal domain (aa 1–162) of SLCV-Lin members (highlighted in red) and its homologous of other geminivirus subgroups. The phylogenetic tree was inferred using the maximum likelihood method based on the general reverse transcriptase model. The percentage of trees in which the associated taxa are clustered together is shown next to the branches (1000 iterations). The red filled circle indicates the node of Rep162 of the SLCV cluster, and the black filled circle indicates the main branch of equivalent Rep162 domains of other geminiviruses. EVS, endogenous viral sequences.

### 3.6. Comparative Analysis of the AC4 Protein of SLCV Lineage Viruses

As the AC4 gene is entirely included in the DNA segment encoding the Rep N-domain, it may be anticipated that AC4 proteins of S-Lin BGVs should be divergent from other geminivirus AC4/C4 proteins. In fact, the alignment of the S-Lin AC4 proteins with their putative equivalents in other geminiviruses revealed a level of amino acid sequence identity so low (<25%) that they would hardly

be considered homologous (i.e., evolutionarily related). In the cladogram illustrated in Figure 5, which was derived from comparisons of AC4/C4 proteins encoded by members from different *Geminiviridae* genera, the S-Lin AC4 proteins form a cluster clearly separated from other geminivirus AC4/C4 proteins. The hypothesis that S-Lin AC4 proteins are not evolutionary related to other AC4/C4 proteins was strongly supported by two distinct observations: (1) the former proteins are significantly larger (~120 aa) than AC4/C4 proteins encoded by “typical” BGVs, topocuviruses, turncurtoviruses, and some curtoviruses, which are 85 aa long on average; (2) S-Lin AC4 proteins lack the *N*-myristoylation motif (MGXLIS) that is critical for geminivirus AC4/C4 function as suppressor of silencing and determinant of pathogenesis [45–47]. Therefore, it is plausible that S-Lin AC4 proteins are neither homologous nor analogous (i.e., with similar functions) to other geminivirus AC4/C4 proteins.



**Figure 5.** Phylogenetic relationships between AC4 proteins encoded by SLCV-Lin members (highlighted in red) and AC4/C4 proteins of other geminivirus subgroups. The phylogenetic tree was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The percentage of trees in which the associated taxa clustered together is shown next to the branches (1000 iterations). Red filled circle indicates the node of the S-Lin BGVs, and the black filled circle indicates the main branch of other geminiviruses.

### 3.7. Comparative Analysis of Curtovirus and S-Lin BGV Genomes

The striking similarity of the N-domain of Rep proteins encoded by S-Lin BGVs and the curtoviruses HrCTV and SpSCTV could be explained by assuming that the former BGVs evolved by a recombination event between a curtovirus and a begomovirus. As the full Rep proteins of SpSCTV and HrCTV display a sequence identity (i.e., 80%) that is practically equal to that showed with the Rep encoded by some S-Lin BGVs (e.g., *Tomato rugose yellow leaf curl virus*; FN434438), we inferred that the hypothetical recombination event plausibly encompassed the entire Rep gene, including the part of C2 ORF that overlaps with the latter gene. Consequently, we search for any evidence of an evolutionary relationship between HrCTV and SpSCTV C2 ORF and AC2/C2 genes of begomoviruses. BLASTp analysis of HrCTV C2 did not produce any significant data. In contrast, computer-assisted analysis of SpSCTV C2 revealed that residues in the protein N-terminus are conserved in AC2/C2 proteins of several Old World BGVs. For instance, the segment (aa 1–12) of SpSCTV C2 protein (i.e., MPYSFPSVNHCT) was 75% identical to the equivalent domain of the AC2 protein of one isolate of *East African cassava mosaic virus* (GenBank accession no. CBA13492), namely, MPsSsPSkNHCT. Moreover, homologous proteins encoded by several BGVs native to Asia (e.g., AYVV and MYMIV), the Mediterranean basin (e.g., TYLCV), and Australia (e.g., ToLCV) display a consensus sequence, MxxSxPSxNHCT, at their N-terminus. The real homology of SpSCTV C2 and begomovirus AC2/C2 genes was confirmed by *in silico* translation (in different frames) of the DNA segment where C1 and C2 ORFs overlap. Indeed, the translation of the 5' end of SpSCTV C2 ORF in the frame of C1 ORF produced the sequence AVFISISEPLY, which is 82% identical to the Rep segment 330–340 of *Sri Lankan cassava mosaic virus* (AtFISISEPLY), 332–342 of *Jatropha mosaic Nigeria virus* (AtFItISEPLY), 331–341 of *Tomato leaf curl China virus* (AVFvSItEPLY), and 335–345 of *Asystasia mosaic Madagascar virus* (AtFItISEPLY), among others. Taken together, the results from this systematic sequence analysis did not provide support for the notion that S-Lin BGVs acquired their distinctive genomic module from a curtovirus; on the contrary, the analysis data strongly support the hypothesis that HrCTV and SpSCTV got its atypical Rep and C4 genes through recombination with an S-Lin begomovirus.

### 3.8. Searching for the Common Ancestor of the SLCV Clade Members

Having ruled out the possibility that a curtovirus related to SpSCTV was the evolutionary ancestor that transferred the atypical genomic module to S-Lin BGVs, we proceeded to search for endogenous viral sequences (EVS) in higher plant genomes, hoping to find some molecular “fossil” with significant similarity with S-Lin BGV sequences. Geminivirus-related EVS have been described in *Nicotiana tomentosiformis* and its relatives [48,49], and in several species of the monocot *Dioscorea* sp. [50]. In addition, the survey of plant genome databases with a variety of bioinformatics tools allowed us to find out Rep-related proteins encoded in the genomes of *Lactuca sativa*, *Corchorus olitorius*, *Coffea canephora*, and *C. arabica*. A close examination of those molecular fossils did not reveal sequence motifs similar to those of S-Lin Rep; in actual fact, several of those EVS-encoded Reps exhibit motifs alike those of “typical” begomoviruses (Figure 6). A systematic analysis of multiple sequence alignments, including hundreds of predicted Rep proteins, disclosed the existence of two major groups of geminivirus Reps, based on the presence of one specific conserved sequence in the 116–134 domain. The distinctive signature of proteins encoded by members of four genera and three unclassified geminiviruses was the conserved consensus FQIDGRSARGGQQ(S/T)AND, of indeterminate function. Its counterpart in Rep proteins, encoded by S-Lin BGVs and the curtoviruses HrCTV and SpSCTV, was a 18 aa long motif, QYKVS GGTKANKDDVYHN (Figure 6). Those two distinctive aa sequences are close to the C-end of the conserved RCR Motif III, which contains the endonuclease active site of geminivirus Rep and a plethora of virus-encoded initiators of rolling-circle replication [51,52].

		RCR Motif III		
AbGMV	*	96-AKDTNAVKN	YITKEGD	YCESGQYKVPG---GTKANKDDVYHNAVN-137
CarYSV	*	96-AKDTNAVRN	YITKEGD	YCESGQYKVSF---GSKANKDDVYHNVVN-137
JacMYuV	*	96-AKDTNAVKN	YITKEGD	YCESGQYKVSF---ATRTNKEDVYHNALH-137
ViYMV	*	96-AKDSNAVKN	YITKEGD	YCESGQYKVSF---ATKSNKDDVYHNAIN-137
SLCV		96-AKDTNAVKN	YITKEGD	YCESGQYKVSF---GSKSNKDDVYHNAVN-137
CbLCV		96-AKDTNAVKN	YITKDG	YCESGQYKVSF---GTKANKDDVYHNAVN-137
RhMYuV		96-AKDTNAVKN	YITKDG	YCESGQYKVSF---GTKANKDDVYHNAVI-137
ToSLCV		96-AKDTNAVQN	YITKEGD	YCESGQYKVSF---GTKSNKDDVYHNAVN-137
JacYVV		96-AKDTNAVKN	YITKEGD	YCESGLYKVSF---GTKSNKDDVYHNAVN-137
TRYLCV		96-AKDTNAVKN	YITKEGD	YCESGQYKVSF---GTKANKDDVYHNAVN-137
ToCmMV		96-AKDTNAVKN	YITKEGD	YCESGQYKVSF---GTKANKDDVYHNAVN-137
SpSCTV		96-AKDANAVHN	YITKEGD	YCESGQYKVSF---SSKANKDDVYHNAVN-137
HrCTV		96-AKDTNAVKN	YITKEGD	YCESGQYKVSF---GTKANKDAVYHNAVN-137
AbMV		94-AKSSSDVKS	YIDKDG	TAEWGEFQIDGRSARGGQQTANDSYAKALN-139
PHYVV		94-AKSSSDVKT	YIDKDG	TVEWGEFQIDGRSARGGQQSANDTYAKALN-139
JacYMV		94-AKSSSDVKS	YLDKDG	VIEWGEFQIDGRSARGGQQTANDTYAKALN-139
BGYMV		94-AKSSSDVKE	YIDKDV	TIEWGQFQVDGRSARGGQQSANDSYAKALN-139
ACMV		94-AKSSSDVKS	YLDKDG	TVEWGQFQIDGRSARGGQQSANDAYAKALN-139
WmCSV		93-AKSSTDVKS	YLDKDG	IVEWGKFQIDGRSARGGQQTANDAYATALN-138
ICMV		94-AKSSSDVKS	YIDKDG	TLEWGTFFQIDGRSARGGQQSANDAYAAALN-139
SwLCV		94-AKSSSDVKS	YVDKDG	TITWGEFQVDGRSARGGQQTANDAAAEALN-139
BCTV-Cal		94-AKSSSDVKS	YVDKDG	TIEWGEFQIDGRSARGGQQTANDSYAKALN-139
BCTV-Mx		94-AKSASDVKQ	YIEKDG	YVDWGTFFQIDGRSARGGQQTANDAAAEALN-139
TCTV		95-AKSSSDVKS	YIEKDG	YIDWGEFQIDARSARGGQQTANDACAEALN-140
SesCTV		95-AKSSTDVKS	YIEKDG	YLDWGEFQIDGRSARGGQQTANDACAEALN-140
TLRoV		95-AKSSSDVKT	YIEKEGE	YSDWGEFQIDGRSSRGGQQSANDAYAAAIN-140
TPCTV		95-AKSSSDVNS	YVDKDG	TIEWGEFQIDARSARGGQQTANDECAEALN-140
GraGV-A		95-AKSSSDVKS	YIEKDG	VISWGEFQIDGRSARGGVQSANDAYAEALN-140
AGmV		97-AKSASDVKT	YIEKGGE	FLDHGVFQIDARSARGGQGHAEVYAEALN-142
JmaV		94-AKSSSDVKN	YIEKDGQ	FADWGEFQIDGRSARGGQQSANDAYASAIN-139
EVS-Dioscorea		95-AKSCSDVKA	YVEKDG	YTDWGEFQIDGRSSRGGTHDLLTRYADASN-140
EVS-Nicotiana		94-AKSSSDVKT	YVEKDG	FIDFGVFQIDGRSSRGGQQSANDSYAKAIN-139
EVS-Corchorus		92-CRSSSDAKE	YIAKDN	FCEWGEFQIDGRSARGGRQSVQDAYAQAIN-137

**Figure 6.** Alignment of the RCR motif III-containing region of Rep proteins encoded by S-Lin BGVs; other NW and OW begomoviruses; members of the genera *Curtoivirus*, *Turncurtoivirus*, and *Topocuvirus*; three unclassified geminiviruses; and several endogenous viral sequences (EVS). The distinctive signature of 17–18 aa residues distinguishing two major evolutionary lineages of geminiviruses is highlighted in light magenta and yellow color, respectively. The name and GenBank accession numbers of the geminiviruses are listed in Supplementary Table S2.

In view of the fact that the Rep gene segment encoding the protein N-domain overlaps with the AC4 gene, we look for significant correlations between the distinctive Rep 116–134 domain and the type of AC4/C4 protein. In all examined cases, viruses encoding Rep proteins with a QYKVSFGTKANKDDVYHN consensus also encode large AC4/C4 proteins lacking a N-myristoylation motif, whereas viruses having Rep with a FQIDGRSARGGQ(S/T)AND sequence usually encode AC4/C4 proteins with a N-myristoylation motif (Figure 7). This rule holds even in the case of Rep and AC4/C4 proteins encoded by EVS integrated in plant genomes, as we were able to know by reconstructing the encoded proteins from the EVS nucleotide sequences (Figures 6 and 7).



**Figure 7.** Alignment of the N-terminal domain of AC4/C4 proteins encoded by members of the S-Lin; other NW and OW begomoviruses; members of the genera *Curtovirus*, *Turncurtovirus*, and *Topocuvirus*; four unclassified geminiviruses; and a number of endogenous viral sequences (EVS). The amino acid residues identical or similar are shadowed in yellow or blue to highlight the resemblances between the aligned protein sequences. The glycine residue of the N-myristoylation motif is marked in red. The name and GenBank accession numbers of the geminiviruses and EVS are enumerated in Supplementary Table S3.

#### 4. Discussion

It is generally accepted that the begomoviruses evolved in the Cretaceous period, prior to the tectonic breakup of the supercontinent Gondwana into the landmasses of India, Africa, South America, Antarctica, and Australia [13,19,27]. This supposition is based on the observation that phylogenies generated from full-length DNA-A alignments consistently produce continent-specific lineages [12,13]. The NW BGVs form a separate major branch, with several distinct clades that are not restricted to specific geographical areas of the Americas. The prevalent view on the origin of NW BGVs is that they evolved from Asian ancestors that arrived to North America probably ~30 MYA, according to the most comprehensive study on this subject that has been performed at present [28]. However, an alternative scenario, that is, that the BGVs arrived to the New World by South America when land connections with a warm Antarctic continent still existed, is likewise supported by several molecular, biogeographical, and paleovirological lines of evidence [53]. Regardless of which of the two alternative scenarios may be true, the evolutionary radiation of BGVs took place during its spread across the two Americas, and eventually some secondary lineages emerged, such as the so-called clades of the SLCV, AbMV, and PHYVV; as well as others identified in diverse phylogenetic studies [27,54]. In this work, we have established, through the genome characterization of four novel BGV species and comprehensive analyses based on a variety of comparative approaches, that the SLCV clade members are quite different from other NW and OW BGVs in a number of molecular traits, including

the replication origin region, and the genes encoding the AC4 protein and the Rep N-domain. Because all those elements are comprised in a compact, unbroken genome module ~670 nt long, it is feasible to analyze it as an evolutionary unit. Thus, the significant divergence of the DNA-binding domain of the S-Lin Rep [24] correlates with the peculiar characteristics of their iterons, which differ from those of other BGVs in two aspects: (a) its number and arrangement within the *Ori* region; (b) its nucleotide sequence, which exhibits a highly conserved consensus, TGGWG(T/a)CC, clearly different to the highly variable aatyGGNRNN iteron consensus of typical NW and OW BGVs [23,24,30]. Likewise, the atypical AC4 protein encoded by the S-Lin BGVs, which lacks a *N*-myristoylation motif, is inseparable from the peculiar Rep N-domain because the corresponding ORFs are overlapped.

In addition to its bizarre iterons, the CR of S-Lin members generally lacks the G-box adjacent to the *Ori* stem-loop element present in most NW BGVs, which is a critical activating element of TGMV *Rep* promoter [55]. Instead, most S-Lin BGVs exhibit a conserved GGGGCAAAA motif contiguous to the *Ori* stem-loop [30]. To ascertain the probable function of this conserved (G4CA4) box, a search in databases specialized in plant *cis*-regulatory elements was carried out; in two of those databases, SoftBerry [56] and PLACE [57], the sequence GGGGCAAAA was identified as a binding-site for the transcription factor E2F of *N. tabacum* [58] and *Arabidopsis thaliana* [59]. Importantly, E2F transcription factors are key components of the molecular machinery responsible for plant cell cycle progression [58,59]. Furthermore, the position of the putative E2F binding site in the common region of S-Lin BGVs is consistent with the general occurrence of *cis*-regulatory elements in the immediate vicinity of the conserved “hairpin” sequence of geminivirus replication origin [23,32,55,60].

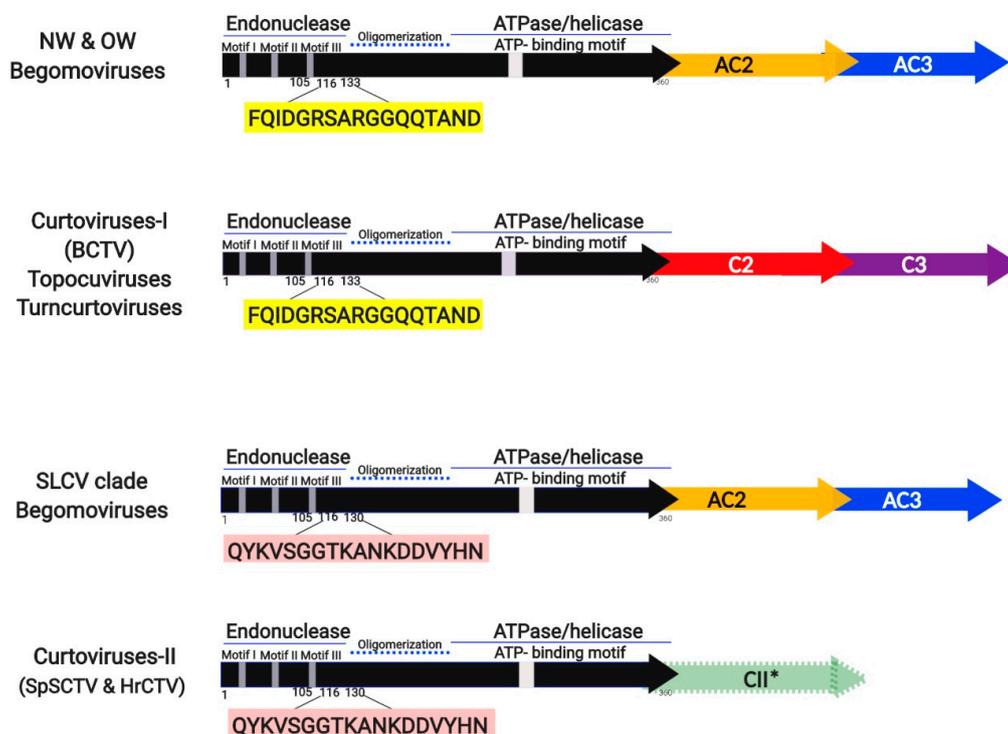
#### *Evolutionary Relationships of Curtoviruses with S-Lin Begomoviruses*

The genus *Curtovirus* comprises four recognized species: BCTV, PepYDV, HrCTV, and SpSCTV; all of them native to USA and México [5,61,62]. The extraordinary resemblance of Rep and AC4 proteins of S-Lin BGVs with the homologous proteins of HrCTV and SpSCTV led to supposing that those BGVs evolved by recombination with a curtovirus. Nevertheless, this hypothesis does not explain several incongruous facts: (1) the Rep proteins of HrCTV and SpSCTV are very divergent from those of BCTV and PepYDV, displaying a sequence identity lower than 45% along its N-terminal (1–160) domain; (2) the *C4* gene of HrCTV and SpSCTV encodes a protein similar to that of S-Lin BGVs, lacking the *N*-myristoylation motif present in the *C4* proteins of BCTV and PepYDV; (3) the *C2* ORF of HrCTV and SpSCTV encodes a ~135 aa protein with no apparent homology (i.e., less than 15% of aa sequence identity) with *C2* proteins of BCTV and PepYDV. Moreover, even though SpSCTV and HrCTV have a *C1* ORF encoding highly similar Rep proteins (SI = 80%), their predicted *C2* proteins display very low sequence identity (26%). In this work, it was demonstrated that HrCTV and SpSCTV are actually recombinants of a curtovirus with a JacMYuV-like begomovirus, according to the phylogenetic relationships of their *C4* and Rep160 proteins depicted in Figure 4 and Figure S6. The gathered evidence supporting this hypothesis is convincing because the *C1/Rep* genes of HrCTV and SpSCTV are definitely homologous to S-Lin BGV *Rep* gene, and the SpSCTV DNA region where the *C1* and *C2* ORFs overlap could be translated *in silico* to a short aa sequence highly similar to the N-terminus of AC2 proteins encoded by several BGVs, and to the 330–340 domain of Rep proteins encoded by the overlapping *C1* gene. Two important corollaries of the notion that HrCTV and SpSCTV derived from a common, recombinant ancestor, are the following: (1) the genus *Curtovirus* is not a monophyletic taxon, and (2) the *C2* ORFs of SpSCTV and HrCTV have undergone an evolutionary divergence that is by far higher than that in the overlapped *C1* ORF, which was jointly acquired by recombination. Both conclusions are important and should be further examined.

On the other hand, the curtoviruses BCTV and PepYDV possess *C2* and *C3* genes that are clearly homologous to those of begomoviruses. The notion that the *C2*–*C3* genes were acquired by those curtoviruses through recombination with a begomovirus has been broadly documented [13,62]. Considering the native geographic distribution of the curtovirus and its vector insect, the leafhopper *Circulifer tenellus*, it is natural to suppose that the intermolecular recombination events that

independently gave rise to BCTV-PepYDV and HrCTV-SpSCTV were possible once the populations of the insect vectors, *Bemisia tabaci* and *Circulifer tenellus*, converged in the same geographical area (i.e., North America). Begomoviruses probably existed in South America before the formation of Panama Isthmus (~3 MYA) because the genomes of *Nicotiana tabacum* and its South American relatives contain endogenous begomoviral sequences that presumably integrated at plant genomes 3–6 MYA [28,49]. Consequently, it is reasonable to assume that modern curtoviruses evolved after the above mentioned geological event, which enabled plant and animal species to migrate between the two Americas by creating an intercontinental land bridge. Similarly, *Tomato pseudocurly top virus*, native to Florida and the only member of the genus *Topocuvirus*, has C1–4 genes of apparent begomoviral origin [63], and probably evolved after the formation of Panama Isthmus.

The data derived of the extensive comparisons of Rep and AC4/C4 proteins performed in this study provide a useful background to think about the complex evolutionary history of the geminiviruses. In Figure 8, a simplified, schematic representation of some our findings is illustrated.



**Figure 8.** Simplified representation of complementary sense genes of begomoviruses and other subgroups of geminiviruses. The characteristic amino acid sequence signature in the aa112–130 domain of Rep proteins uncovered in this study is highlighted in color boxes. AC2 and AC3 genes of begomoviruses are represented by arrows in yellow and blue color, respectively. The C2 and C3 genes of members of other *Geminiviridae* genera are differently colored to emphasize structural and/or functional divergence from its begomoviral homologous. The green arrow of SpSCTV and HrCTV curtoviruses represents the highly divergent C2 (i.e., CII\*) open reading frame (ORF), of unknown function. NW, New World; OW, Old World.

## 5. Conclusions

The members of the SLCV clade (35 species) encode unique AC4 and Rep proteins whose evolutionary origin is enigmatic. The actual function of that atypical AC4 protein is unknown, but probably differs from those of other AC4/C4 proteins of begomoviruses and curtoviruses. The replication origin region of the S-Lin BGVs harbors peculiar iterons and a putative binding site for E2F transcription factors, involved in the plant cell cycle, which is adjacent to the conserved stem-loop sequence. A virus basal to the SLCV lineage, JacMYuV, which is the closest relative to the hypothetical begomoviral

ancestor of two recombinant curtoviruses, was characterized in this work. With the exception of the S-Lin BGVs, most geminiviruses displaying a Rep gene without introns encode proteins with a FQIDGRSARGGQQ(S/T)AND related consensus near the C-end of the endonuclease active site (i.e., the RCR Motif III). The actual function of the former 17aa long motif sited between the endonuclease and the oligomerization domains of Rep is unclear, although a study of TGMV Rep protein showed that specific mutations in the so-called “conserved sequence” (i.e., FQVDGRSARGGCQT) impaired several key Rep activities, including DNA-binding and cleavage, and protein oligomerization [64]. These observations suggest that the Motif III-associated conserved sequence is a critical subdomain of begomovirus replication proteins.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1999-4915/11/7/644/s1>, Figure S1: Weeds infected with the begomoviruses described in this study, Figure S2: Phylogenetic tree and matrix of pairwise sequence identity based on DNA-A sequences of the isolated begomoviruses and their relatives, Figure S3: Recombination event in the AbGMYuV DNA-A detected by RDP, Figure S4: Phylogenetic relationships between the novel begomoviruses and selected New World and Old World begomoviruses, Figure S5: DNA-B inferred evolutionary relationships of viruses isolated in this work compared with selected begomoviruses. Figure S6: Phylogenetic relationships between AC4 proteins encoded by SLCV-Lin members (highlighted in red) and those of other geminivirus subgroups, Table S1: Oligonucleotides used in this work, Table S2: Names, acronyms and GenBank accession numbers of geminiviruses and endogenous viral sequences (EVS) included in the alignments of Rep motif III region and the N-terminal end of AC4 proteins (Figure 6), Table S3: Names, acronyms and GenBank accession numbers of geminiviruses and endogenous viral sequences (EVS) included in the alignments of the N-terminal domain of AC4/C4 proteins (Figure 7).

**Author Contributions:** G.R.A.-A. and A.R.-O. conceived the research project; O.M.-V. and Y.M.-G. carried out sample collection; S.I.T.-H. and S.A.-G. isolated and characterized the new viruses; S.I.T.-H. carried out the plant infection experiments; Y.C.-C. and G.P.-P. carried out the recombination analysis; Y.C.-C., J.H.R.-P., and L.R.-R. performed the phylogenetic analyses; A.R.-O. and G.R.A.-A. carried out the comparative genome analysis of geminiviruses; A.R.-O., Y.C.-C., and G.P.P. prepared the figures; G.R.A.-A. administered the project; G.R.A.-A. and A.R.-O. wrote the original draft; G.R.A.-A. revised and edited the final manuscript. All authors reviewed and approved the manuscript.

**Funding:** This research was supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT) of Mexico (URL—<https://www.conacyt.gob.mx/>). Grant SEP-CB-2015-257136 to G.R.A.-A.

**Acknowledgments:** We thank Isabel Isordia Jasso for skilful assistance in the technical work, and Bernardo Bañuelos-Hernández and Aarón Avalos-Calleros for critical reading of the manuscript and thoughtful suggestions. We would like to thank the National Laboratory LANBAMA for the technical assistance provided. ARO and SITH were supported by a fellowship from the Consejo Nacional de Ciencia y Tecnología (CONACYT, Mexico).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hanley-Bowdoin, L.; Bejarano, E.R.; Robertson, D.; Mansoor, S. Geminiviruses: Masters at redirecting and reprogramming plant processes. *Nat. Rev. Microbiol.* **2013**, *11*, 777. [[CrossRef](#)] [[PubMed](#)]
2. Hesketh, E.L.; Saunders, K.; Fisher, C.; Potze, J.; Stanley, J.; Lomonosoff, G.P.; Ranson, N.A. The 3.3 Å structure of a plant geminivirus using cryo-EM. *Nat. Commun.* **2018**, *9*, 2369. [[CrossRef](#)] [[PubMed](#)]
3. Xu, X.; Zhang, Q.; Hong, J.; Li, Z.; Zhang, X.; Zhou, X. Cryo-EM Structure of a Begomovirus Geminiate Particle. *Int. J. Mol. Sci.* **2019**, *20*, 1738. [[CrossRef](#)] [[PubMed](#)]
4. Rojas, M.R.; Macedo, M.A.; Maliano, M.R.; Soto-Aguilar, M.; Souza, J.O.; Briddon, R.W.; Kenyon, L.; Rivera-Bustamante, R.F.; Zerbini, F.M.; Adkins, S.; et al. World Management of *Geminiviruses*. *Annu. Rev. Phytopathol.* **2018**, *56*, 637–677. [[CrossRef](#)] [[PubMed](#)]
5. Zerbini, F.M.; Briddon, R.W.; Idris, A.; Martin, D.P.; Moriones, E.; Navas-Castillo, J.; Rivera-Bustamante, R.; Rourmagnac, P.; Varsani, A. ICTV virus taxonomy profile: *Geminiviridae*. *J. Gen. Virol.* **2017**, *98*, 131–133. [[CrossRef](#)] [[PubMed](#)]
6. Varma, A.; Malathi, V.G. Emerging geminivirus problems: A serious threat to crop production. *Ann. Appl. Biol.* **2003**, *142*, 145–164. [[CrossRef](#)]
7. Varsani, A.; Navas-Castillo, J.; Moriones, E.; Hernández-Zepeda, C.; Idris, A.; Brown, J.K.; Murilo Zerbini, F.; Martin, D.P. Establishment of three new genera in the family *Geminiviridae*: Becurtovirus, Eragrovirus and Turncurtovirus. *Arch. Virol.* **2014**, *159*, 2193–2203. [[CrossRef](#)]

8. Varsani, A.; Roumagnac, P.; Fuchs, M.; Navas-Castillo, J.; Moriones, E.; Idris, A.; Briddon, R.W.; Rivera-Bustamante, R.; Murilo Zerbini, F.; Martin, D.P. Capulavirus and Grablovirus: Two new genera in the family *Geminiviridae*. *Arch. Virol.* **2017**, *162*, 1819–1831. [[CrossRef](#)]
9. International Committee on Taxonomy of Viruses (ICTV). Available online: <https://talk.ictvonline.org/taxonomy/> (accessed on 8 July 2019).
10. De Barro, P.J.; Liu, S.S.; Boykin, L.M.; Dinsdale, A.B. *Bemisia tabaci*: A statement of species status. *Ann. Rev. Entomol.* **2011**, *56*, 1–19. [[CrossRef](#)]
11. Waqar, I.; Akutse, K.S.; Qasim, M.; Khan, K.A.; Ghramh, H.A.; Idrees, A.; Latif, S. *Bemisia tabaci*-mediated facilitation in diversity of begomoviruses: Evidence from recent molecular studies. *Microb. Pathog.* **2018**, *123*, 162–168. [[CrossRef](#)]
12. Briddon, R.W.; Patil, B.L.; Bagewadi, B.; Nawaz-ul-Rehman, M.S.; Fauquet, C.M. Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. *BMC Evol. Biol.* **2010**, *10*, 97. [[CrossRef](#)] [[PubMed](#)]
13. Rybicki, E.P. A phylogenetic and evolutionary justification for three genera of Geminiviridae. *Arch. Virol.* **1994**, *139*, 49–77. [[CrossRef](#)] [[PubMed](#)]
14. Macedo, M.A.; Albuquerque, L.C.; Maliano, M.R.; Souza, J.O.; Rojas, M.R.; Inoue-Nagata, A.K.; Gilbertson, R.L. Characterization of tomato leaf curl purple vein virus, a new monopartite New World begomovirus infecting tomato in Northeast Brazil. *Arch. Virol.* **2018**, *163*, 737–743. [[CrossRef](#)] [[PubMed](#)]
15. Melgarejo, T.A.; Kon, T.; Rojas, M.R.; Paz-Carrasco, L.; Zerbini, F.M.; Gilbertson, R.L. Characterization of a new world monopartite begomovirus causing leaf curl disease of tomato in Ecuador and Peru reveals a new direction in geminivirus evolution. *J. Virol.* **2013**, *87*, 5397–5413. [[CrossRef](#)] [[PubMed](#)]
16. Romay, G.; Geraud-Pouey, F.; Chirinos, D.T.; Mahillon, M.; Gillis, A.; Mahillon, J.; Bragard, C. Tomato Twisted Leaf Virus: A Novel Indigenous New World Monopartite Begomovirus Infecting Tomato in Venezuela. *Viruses* **2019**, *11*, 327. [[CrossRef](#)] [[PubMed](#)]
17. Brown, J.K.; Zerbini, F.M.; Navas-Castillo, J.; Moriones, E.; Ramos-Sobrinho, R.; Silva, J.C.; Fiallo-Olivé, E.; Briddon, R.W.; Hernández-Zepeda, C.; Idris, A.; et al. Revision of Begomovirus taxonomy based on pairwise sequence comparisons. *Arch. Virol.* **2015**, *160*, 1593–1619. [[CrossRef](#)]
18. Fondong, V.N. Geminivirus protein structure and function. *Mol. Plant Patho.* **2013**, *14*, 635–649. [[CrossRef](#)]
19. Rojas, M.R.; Hagen, C.; Lucas, W.J.; Gilbertson, R.L. Exploiting chinks in the plant's armor: Evolution and emergence of *geminiviruses*. *Annu. Rev. Phytopathol.* **2005**, *43*, 361–394. [[CrossRef](#)]
20. Lazarowitz, S.G.; Beachy, R.N. Viral movement proteins as probes for intracellular and intercellular trafficking in plants. *Plant Cell* **1999**, *11*, 535–548. [[CrossRef](#)]
21. Laufs, J.; Traut, W.; Heyraud, F.; Matzeit, V.; Rogers, S.G.; Schell, J.; Gronenborn, B. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 3879–3883. [[CrossRef](#)]
22. Ha, C.; Coombs, S.; Revill, P.; Harding, R.; Vu, M.; Dale, J. Corchorus yellow vein virus, a New World geminivirus from the Old World. *J. Gen. Virol.* **2006**, *87*, 997–1003. [[CrossRef](#)] [[PubMed](#)]
23. Argüello-Astorga, G.R.; Guevara-Gonzalez, R.G.; Herrera-Estrella, L.R.; Rivera-Bustamante, R.F. Geminivirus replication origins have a group-specific organization of iterative elements: A model for replication. *Virology* **1994**, *203*, 90–100. [[CrossRef](#)] [[PubMed](#)]
24. Argüello-Astorga, G.R.; Ruiz-Medrano, R. An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: Identification of potential interacting amino acid–base pairs by a comparative approach. *Arch. Virol.* **2001**, *146*, 465–485. [[CrossRef](#)]
25. Ha, C.; Coombs, S.; Revill, P.; Harding, R.; Vu, M.; Dale, J. 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. *J. Gen. Virol.* **2008**, *89*, 312–326. [[CrossRef](#)] [[PubMed](#)]
26. Mauricio-Castillo, J.A.; Torres-Herrera, S.I.; Cardenas-Conejo, Y.; Pastor-Palacios, G.; Mendez-Lozano, J.; Argüello-Astorga, G.R. A novel begomovirus isolated from sida contains putative cis- and trans-acting replication specificity determinants that have evolved independently in several geographical lineages. *Arch. Virol.* **2014**, *159*, 2283–2294. [[CrossRef](#)]
27. Nawaz-ul-Rehman, M.S.; Fauquet, C.M. Evolution of geminiviruses and their satellites. *FEBS Lett.* **2009**, *583*, 1825–1832. [[CrossRef](#)] [[PubMed](#)]

28. Lefeuvre, P.; Harkins, G.W.; Lett, J.M.; Briddon, R.W.; Chase, M.W.; Moury, B.; Martin, D.P. Evolutionary timescale of the begomoviruses: Evidence from integrated sequences in the *Nicotiana* genome. *PLoS ONE* **2011**, *6*, e19193. [[CrossRef](#)]
29. Brown, J.K.; Idris, A.M.; Alteri, C.; Stenger, D.C. Emergence of a New Cucurbit-Infecting Begomovirus Species Capable of Forming Viable Reassortants with Related Viruses in the Squash leaf curl virus Cluster. *Phytopathology* **2002**, *92*, 734–742. [[CrossRef](#)]
30. Gregorio-Jorge, J.; Bernal-Alcocer, A.; Bañuelos-Hernández, B.; Alpuche-Solís, A.G.; Hernández-Zepeda, C.; Moreno-Valenzuela, O.; Frías-Treviño, G.; Argüello-Astorga, G.R. Analysis of a new strain of Euphorbia mosaic virus with distinct replication specificity unveils a lineage of begomoviruses with short Rep sequences in the DNA-B intergenic region. *Virol. J.* **2010**, *7*, 275. [[CrossRef](#)]
31. Márquez-Martín, B.; Maeso, D.; Martínez-Ayala, A.; Bernal, R.; Teresa-Federici, M.; Vincelli, P.; Navas-Castillo, J.; Moriones, E. Diverse population of a new bipartite begomovirus infecting tomato crops in Uruguay. *Arch. Virol.* **2012**, *157*, 1137–1142. [[CrossRef](#)]
32. Argüello-Astorga, G.; Herrera-Estrella, L.; Rivera-Bustamante, R. Experimental and theoretical definition of geminivirus origin of replication. *Plant Mol. Biol.* **1994**, *26*, 553–556. [[CrossRef](#)] [[PubMed](#)]
33. Dellaporta, S.L.; Wood, J.; Hicks, J.B. A plant DNA miniprep: Version II. *Plant Mol. Biol. Rep.* **1983**, *1*, 19–21. [[CrossRef](#)]
34. Mauricio-Castillo, J.A.; Argüello-Astorga, G.R.; Ambriz-Granados, S.; Alpuche-Solís, A.G.; Monreal-Vargas, C.T. First Report of Tomato golden mottle virus on *Lycopersicon esculentum* and *Solanum rostratum* in Mexico. *Plant Dis.* **2007**, *91*, 1513. [[CrossRef](#)] [[PubMed](#)]
35. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
36. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [[CrossRef](#)] [[PubMed](#)]
37. Tamura, K.; Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* **1993**, *10*, 512–526. [[CrossRef](#)] [[PubMed](#)]
38. Martin, D.P.; Lemey, P.; Lott, M.; Moulton, V.; Posada, D.; Lefeuvre, P. RDP3: A flexible and fast computer program for analyzing recombination. *Bioinformatics* **2010**, *26*, 2462–2463. [[CrossRef](#)]
39. Hernandez-Zepeda, C.; Idris, A.M.; Carnevali, G.; Brown, J.K.; Moreno-Valenzuela, O.A. Molecular characterization and phylogenetic relationships of two new bipartite begomovirus infecting malvaceous plants in Yucatan, Mexico. *Virus Genes* **2007**, *35*, 369–377. [[CrossRef](#)] [[PubMed](#)]
40. Hernández-Zepeda, C.; Argüello-Astorga, G.R.; Idris, A.M.; Carnevali, G.; Brown, J.K.; Moreno-Valenzuela, O.A. Molecular characterization and phylogenetic relationships of Desmodium leaf distortion virus (DeLDV): A new begomovirus infecting *Desmodium glabrum* in Yucatan, Mexico. *Virus Genes* **2009**, *39*, 371–374. [[CrossRef](#)]
41. Argüello-Astorga, G.R. Department of Molecular Biology, IPICYT, San Luis Potosí, S.L.P., Mexico. Begomovirus phylogeny based on the alignment of the full-length DNA-A. Unpublished work, 2019.
42. Qazi, J.; Ilyas, M.; Mansoor, S.; Briddon, R.W. Legume yellow mosaic viruses: Genetically isolated begomoviruses. *Mol. Plant Pathol.* **2007**, *8*, 343–438. [[CrossRef](#)]
43. Klute, K.A.; Nadler, S.A.; Stenger, D.C. Horseradish curly top virus is a distinct subgroup II geminivirus species with rep and C4 genes derived from a subgroup III ancestor. *J. Gen. Virol.* **1996**, *77*, 1369–1378. [[CrossRef](#)] [[PubMed](#)]
44. Hernandez-Zepeda, C.; Brown, J.K. First report of a new curtovirus species, Spinach severe curly top virus, in commercial spinach plants (*Spinacia oleracea*) from south-central Arizona. *Plant Dis.* **2010**, *94*, 917. [[CrossRef](#)] [[PubMed](#)]
45. Fondong, V.N.; Reddy, R.C.; Lu, C.; Hankoua, B.; Felton, C.; Czymmek, K.; Achenjang, F. The consensus N-myristoylation motif of a geminivirus AC4 protein is required for membrane binding and pathogenicity. *Mol. Plant Microbe Interact.* **2007**, *20*, 380–391. [[CrossRef](#)] [[PubMed](#)]
46. Hipp, K.; Rau, P.; Schäfer, B.; Pfannstiel, J.; Jeske, H. Translation, modification and cellular distribution of two AC4 variants of *African cassava mosaic virus* in yeast and their pathogenic potential in plants. *Virology* **2016**, *498*, 136–148. [[CrossRef](#)] [[PubMed](#)]
47. Li, H.; Zeng, R.; Chen, Z.; Liu, X.; Cao, Z.; Xie, Q.; Yang, C.; Lai, J. S-acylation of a geminivirus C4 protein is essential for regulating the CLAVATA pathway in symptom determination. *J. Exp. Bot.* **2018**, *69*, 4459–4468. [[CrossRef](#)] [[PubMed](#)]

48. Bejarano, E.R.; Khashoggi, A.; Witty, M.; Lichtenstein, C. Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. *Proc. Natl. Acad. Sci. USA* **1996**, *23*, 759–764. [[CrossRef](#)]
49. Murad, L.; Bielawski, J.P.; Matyasek, R.; Kovariik, A.; Nichols, R.A.; Leitch, A.R.; Lichtenstein, C.P. The origin and evolution of geminivirus-related DNA sequences in *Nicotiana*. *Heredity* **2004**, *92*, 352–358. [[CrossRef](#)] [[PubMed](#)]
50. Filloux, D.; Murrell, S.; Koohapitagtam, M.; Golden, M.; Julian, C.; Galzi, S.; Uzest, M.; Rodier-Goud, M.; D’Hont, A.; Vernerey, M.S.; et al. The genomes of many yam species contain transcriptionally active endogenous geminiviral sequences that may be functionally expressed. *Virus Evol.* **2015**, *1*, vev002. [[CrossRef](#)]
51. Chandler, M.; de la Cruz, F.; Dyda, F.; Hickman, A.B.; Moncalian, G.; Ton-Hoang, B. Breaking and joining single-stranded DNA: The HUH endonuclease superfamily. *Nat. Rev. Microbiol.* **2013**, *11*, 525–538. [[CrossRef](#)]
52. Koonin, E.V.; Dolja, V.V.; Krupovic, M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* **2015**, *479–480*, 2–25. [[CrossRef](#)]
53. Cardenas-Conejo, Y. Reconstruction of Genome Evolution of Begomoviruses through an Integrated Multidisciplinary Approach. Ph.D. Thesis, Instituto Potosino de Investigación Científica y Tecnológica (IPICYT), México, 2012. (In Spanish)
54. Fauquet, C.M.; Briddon, R.W.; Brown, J.K.; Moriones, E.; Stanley, J.; Zerbini, M.; Zhou, X. Geminivirus strain demarcation and nomenclature. *Arch. Virol.* **2008**, *153*, 783–821. [[CrossRef](#)] [[PubMed](#)]
55. Orozco, B.M.; Gladfelter, H.J.; Settlege, S.B.; Eagle, P.A.; Gentry, R.N.; Hanley-Bowdoin, L. Multiple cis elements contribute to geminivirus origin function. *Virology* **1998**, *242*, 346–356. [[CrossRef](#)] [[PubMed](#)]
56. SoftBerry. Available online: <http://www.softberry.com/> (accessed on 18 June 2019).
57. Higo, K.; Ugawa, Y.; Iwamoto, M.; Korenaga, T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* **1999**, *27*, 297–300. [[CrossRef](#)] [[PubMed](#)]
58. Chabouté, M.E.; Clément, B.; Sekine, M.; Philipps, G.; Chaubet-Gigota, N. Cell Cycle Regulation of the Tobacco Ribonucleotide Reductase Small Subunit Gene Is Mediated by E2F-like Elements. *Plant Cell.* **2000**, *12*, 1987–1999. [[CrossRef](#)] [[PubMed](#)]
59. Vandepoele, K.; Vlieghe, K.; Florquin, K.; Hennig, L.; Beemster, G.T.; Gruissem, W.; van de Peer, Y.; Inzé, D.; de Veylder, L. Genome-wide Identification of potential plant E2F target genes. *Plant Physiol.* **2005**, *139*, 316–328. [[CrossRef](#)] [[PubMed](#)]
60. Fenoll, C.; Schwarz, J.J.; Black, D.M.; Schneider, M.; Howell, S.H. The intergenic region of maize streak virus contains a GC-rich element that activates rightward transcription and binds maize nuclear factors. *Plant Mol. Biol.* **1990**, *15*, 865–877. [[CrossRef](#)] [[PubMed](#)]
61. Claverie, S.; Bernardo, P.; Kraberger, S.; Hartnadyk, P.; Lefeuvre, P.; Lett, J.M.; Galzi, S.; Filloux, D.; Harkins, G.W.; Varsani, A.; et al. From Spatial Metagenomics to Molecular Characterization of Plant Viruses: A Geminivirus Case Study. *Adv. Virus Res.* **2018**, *101*, 55–83. [[PubMed](#)]
62. Varsani, A.; Martin, D.P.; Navas-Castillo, J.; Moriones, E.; Hernández-Zepeda, C.; Idris, A.; Murilo Zerbini, F.; Brown, J.K. Revisiting the classification of curtoviruses based on genome-wide pairwise identity. *Arch. Virol.* **2014**, *159*, 1873–1882. [[CrossRef](#)]
63. Briddon, R.W.; Bedford, I.D.; Tsai, J.H.; Markham, P.G. Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. *Virology* **1996**, *219*, 387–394. [[CrossRef](#)]
64. Orozco, B.M.; Kong, L.-J.; Batts, L.A.; Elledge, S.; Hanley-Bowdoin, L. The multifunctional character of a geminivirus replication protein is reflected by its complex oligomerization properties. *J. Biol. Chem.* **2000**, *275*, 6114–6122. [[CrossRef](#)] [[PubMed](#)]

