



Article

Gene Overlapping as a Modulator of *Begomovirus* Evolution

Iván Martín-Hernández ^{1,†} and Israel Pagán ^{1,2,*} ¹ Centro de Biotecnología y Genómica de Plantas UPM-INIA, 28223 Madrid, Spain; imartin@iqfr.csic.es² Departamento de Biotecnología—Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, 28045 Madrid, Spain

* Correspondence: jesuisrael.pagan@upm.es; Tel.: +34-91-067-91-65

† Current address: Department of Biological Chemical Physics, Rocasolano Institute of Physical Chemistry CSIC, 28006 Madrid, Spain.

Abstract: In RNA viruses, which have high mutation—and fast evolutionary—rates, gene overlapping (i.e., genomic regions that encode more than one protein) is a major factor controlling mutational load and therefore the virus evolvability. Although DNA viruses use host high-fidelity polymerases for their replication, and therefore should have lower mutation rates, it has been shown that some of them have evolutionary rates comparable to those of RNA viruses. Notably, these viruses have large proportions of their genes with at least one overlapping instance. Hence, gene overlapping could be a modulator of virus evolution beyond the RNA world. To test this hypothesis, we use the genus *Begomovirus* of plant viruses as a model. Through comparative genomic approaches, we show that terminal gene overlapping decreases the rate of virus evolution, which is associated with lower frequency of both synonymous and nonsynonymous mutations. In contrast, terminal overlapping has little effect on the pace of virus evolution. Overall, our analyses support a role for gene overlapping in the evolution of begomoviruses and provide novel information on the factors that shape their genetic diversity.

Keywords: overlapping genes; rate of evolution; begomoviruses; ssDNA viruses



Citation: Martín-Hernández, I.; Pagán, I. Gene Overlapping as a Modulator of *Begomovirus* Evolution. *Microorganisms* **2022**, *10*, 366. <https://doi.org/10.3390/microorganisms10020366>

Academic Editor: Elvira Fiallo-Olivé

Received: 17 December 2021

Accepted: 1 February 2022

Published: 4 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Genomic regions that encode more than one protein, that is, gene overlapping, are commonplace among viruses [1,2]. Such regions have important biological and evolutionary implications. First, they are associated with virus within-host multiplication, between-host transmission, disease severity and strength of host immune response [3–6]. Second, viruses are subjected to strong selection for maintaining smaller genomes because this (i) reduces the chances for deleterious mutations to become fixed in the virus genome, particularly in viruses with high mutation rates; (ii) improves virus fitness due to faster replication; and (iii) optimizes virion formation due to physical limitations imposed by the capsid size [7–9]. Gene overlapping allows increasing the amount of genomic information in viral genomes while controlling for limited capsid space and speeding up the purification of deleterious mutations from the virus population by amplifying their effect, as in overlapping regions these mutations affect more than one gene at the same time [1,9,10].

If gene overlapping is selectively advantageous for viruses, it would be expected to be more frequent: in RNA than in DNA viruses, as the former have (in general) higher mutation rates [11]; in larger than in shorter viral genomes to minimize the chances of deleterious mutations to become fixed [12], and in spherical virions as these generally have smaller inner volumes than other capsid shapes [13]. Although Brades and Linial [9] failed to detect an association between virion shape and frequency of gene overlapping in support of the predictions above, it has been shown that the larger the gene overlapping the greater the reduction in the rate of RNA virus evolution [1], and that gene overlapping appears to

be more frequent in DNA viruses, which on average have also larger genome sizes [11], than in RNA viruses [2].

The species of the family *Geminiviridae* of plant viruses are notable exceptions to the virus characteristics associated with gene overlapping. Although geminiviruses are ssDNA viruses, and therefore replicate through high fidelity polymerases [14], and have small genomes, these viruses have a large proportion of their genes with at least one overlapping region [2]. The *Geminiviridae* family is currently divided into nine genera of which the largest one is the genus *Begomovirus*. This is also one of the most numerous genera of plant viruses with more than 400 species [15]. Bipartite begomovirus genomes encode six open reading frames (ORFs): two in the virion strand (AV1 and AV2) and four in the complementary strand (AC1, AC2, AC3 and AC4), with monopartite begomoviruses encoding equivalent proteins with the same names but without the A prefix. The AV1 gene encodes the coat protein (CP), which is essential for genome encapsidation, viral movement and insect transmission. The AV2 gene, which is considered as the pathogenicity gene, is also involved in movement and symptom development, and functions as a suppressor of gene silencing. This ORF is not present in New World bipartite begomoviruses. Viral DNA replication depends on the AC1 gene product (replication initiator protein, Rep). The AC2 gene encodes the transcriptional activator protein (TrAP) that interferes with transcriptional and post-transcriptional gene silencing (TGS and PTGS, respectively), and with the CP expression. The gene encoding for the AC3 protein (replication enhancer protein, REn), enhances viral DNA accumulation, and is involved in interaction with the plant-host retinoblastoma-related (RBR) proteins. Finally, AC4 counteracts PTGS by inhibiting accumulation of siRNA and is considered an important symptom determinant [15]. All of these six ORFs have at least one overlapping region in both mono- and bipartite begomoviruses [15]. Bipartite begomoviruses have two additional non-overlapping ORFs in the B-component: BC1, a movement protein (MP), and BV1, the nuclear shuttle protein (NSP) (Figure 1) [16].

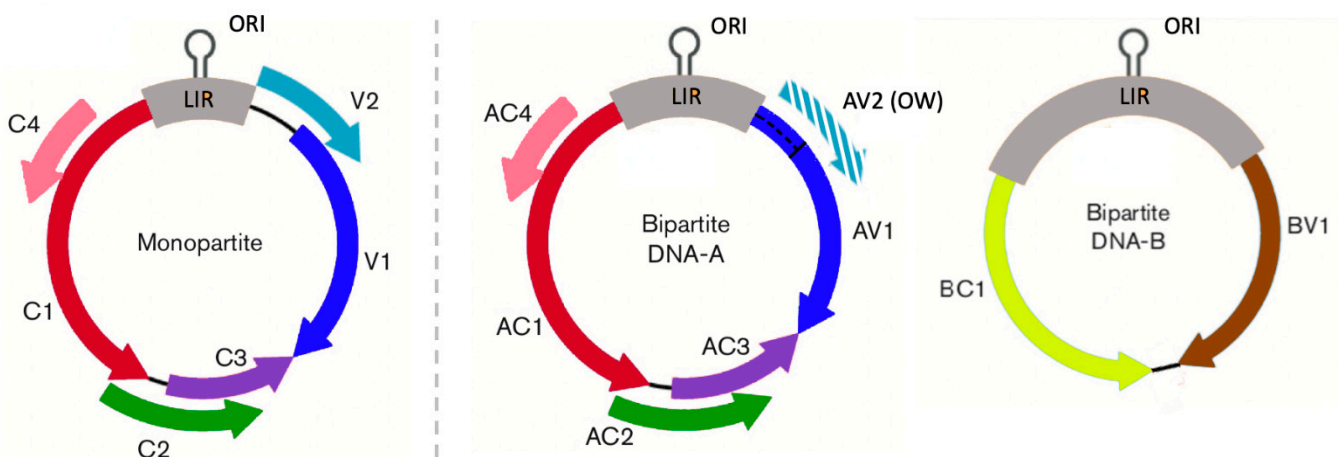


Figure 1. Genome organization of mono- and bipartite begomoviruses. Colored arrows denote the position and orientation of each gene. Monopartite and DNA-A of bipartite begomoviruses encode for: AC1: Replication initiator protein (Rep); AC2: Transcriptional activator protein (TrAP); AC3: Replication enhancer protein (Ren); AC4: Silencing suppressor; AV1: Coat protein (CP); and AV2: Various functions. In bipartite begomoviruses AV2 is only present in Old World (OW) begomoviruses, where AV1 is as long as in monopartite begomoviruses (dashed). DNA-B of bipartite begomoviruses encodes for: BC1: Movement protein (MP) and BC2: Nuclear shuttle protein (NSP). CR, common region. The hairpin which includes the origin of replication (ORI) is indicated in the Long Intergenic Region (LIR) (modified from [17]).

Despite being DNA viruses, begomoviruses have been repeatedly shown to have high evolutionary rates (reviewed by [18]). For instance, *Tomato yellow leaf curl virus* (TYLCV)

substitution rate has been estimated to be of 2.88×10^{-4} nucleotide substitutions per site per year, which is in the range of values for RNA viruses [12]. This fast evolutionary rate has been attributed to the effect of oxidative damage in replicated viral genomes, and/or to higher mutation rates than expected for DNA viruses [12]. If so, extensive gene overlapping in begomoviruses may contribute to modulate mutational load, and consequently the rate of virus evolution, as it has been shown for RNA viruses [1]. Experimental evidence supporting this idea is scarce and sometimes contradictory. For instance, a higher variability occurred in the *Tomato yellow leaf curl China virus* (TYLCCNV) AC1-AC4 overlapping (OV) region than in the non-overlapping (NOV) region of AC1 [19], whereas the opposite was observed for *Pepper huasteco yellow vein virus* (PHYVV) [20].

Here, we analyzed the effect of gene overlapping on the rate of begomovirus evolution through comparative genomics and utilizing sequences from 18 species. In particular, we explored whether the following evolutionary parameters vary between OV and NOV regions and among different types of gene overlap: (1) the rate of viral evolution, using overall tree length as a proxy, (2) the frequency of synonymous and nonsynonymous substitutions, (3) selection pressure and (4) magnitude of the effect of gene overlapping in the rate of virus evolution.

2. Materials and Methods

2.1. Sequence Data

Available sequences from begomovirus species were retrieved from GenBank. Sequences from extensively passaged isolates in non-natural hosts were excluded. When possible, we tried to minimize the presence of recombination. Species with more than 10 sequences were retained for analysis, so that we were able to include 18 mono and bipartite begomoviruses, and a total of 8239 sequences. Overall, we analyzed 125 instances of gene overlap ranging between 59 and 423 nt in length: 17 internal overlapping instances, 54 5'-terminal overlapping instances, and 54 3'-terminal overlapping instances. For simplicity, genes are named as for bipartite begomoviruses. Note that we divided sequences from *Bhendi yellow vein mosaic virus* (BYVMV) into two groups: one of sequences originally classified as belonging to this virus, and another originally characterized as *Bhendi yellow vein India virus*. Although both groups are currently considered as belonging to BYVMV [15], we chose to analyze them separately as evolutionary parameters differed between groups (Appendix A). However, analyses merging the two groups did not change our conclusions. We constructed sequence alignments for the 125 overlapping instances, and for the corresponding OV and NOV fragments of each gene. Sequence alignments of the OV regions were adjusted according to the amino acid sequence of each of the two genes involved, thus generating two data sets for each OV region. All alignments were built using MUSCLE 3.7 [21] and adjusted manually according to the amino acid sequences using AliView [22]. Alignments are available as Supplementary Material File S1.

2.2. Estimation of Tree Length

Tree lengths (t) were estimated for the OV and NOV regions of each gene. To do so, we used a maximum likelihood fitting of the General Time Reversible (GTR) nucleotide substitution model as implemented in the HyPhy package [23]. Differences in total tree length between OV and NOV regions were analyzed using a relative ratio test also utilizing HyPhy. Because t is dependent on the number of tree branches (i.e., number of sequences), when values were compared among overlapping instances, t was normalized according to the number of sequences.

2.3. Selection Pressures

Selection pressures for OV and NOV regions were estimated as the difference between the mean number of nonsynonymous (d_N) and synonymous (d_S) nucleotide substitutions per site (d_N/d_S) using the fast unbiased Bayesian approximation (FUBAR), and the fixed effect likelihood (FEL) methods implemented in HyPhy (Appendix A). Because the two

methods yielded similar results, only the FUBAR results are shown here. In all cases, d_N/d_S measures were based on neighbor-joining trees inferred using the MG94 nucleotide substitution model. Significant differences between d_N/d_S values in OV and NOV regions, were analyzed using a population level adaptation test [24]. Values of d_N and d_S were also estimated. For each pair of overlapping genes, d_N , d_S , and d_N/d_S estimates were obtained for the two reading frames of the OV region. To do so, we used separated sequence alignments for the two overlapping genes, and we partitioned codons, such that OV and NOV regions could be defined over the full-length sequence of each gene.

2.4. Detection of Recombination

For each pair of overlapping genes, recombination breakpoints were detected using six different methods as implemented in RDP5: RDP, GENECONV, MaxChi, 3Seq, Bootscan, and Chimaera [25]. Only recombination signals detected by at least four methods ($p < 0.05$) were considered as positive. For the purpose of this work, recombinants with breakpoints in the LIR and the V1/C3 limit, which are recombination hotspots [26], were not counted as such as they were not differentially affecting OV and NOV regions of any given gene. Instances with more than 10% of recombinant sequences, regardless of breakpoints were located in OV or NOV regions, were considered to have excessive recombination (Appendix A). Analyses were repeated excluding such instances, but conclusions did not vary. Hence, we present here results obtained using all instances.

2.5. Statistical Analysis

The 125 overlapping instances were used for statistical analysis. Tree lengths (t) were not homoscedastic according to Kolmogorov–Smirnov and Levene’s tests. Therefore, this variable was fitted to a gamma distribution; whereas the ratio OV/NOV for t , d_N , d_S , and d_N/d_S , and percentage of overlapping were fitted to a normal distribution, according to Akaike’s Information Criteria (R package: RRISKDISTRIBUTIONS; [27]). Consequently, differences in values of these variables between OV and NOV regions and between types of gene overlap were analyzed by generalized linear models (GzLM), using type of region or type of overlapping as factors. Differences in the proportion of genes for which parameters above differed between OV and NOV regions was analyzed by Fisher’s exact test [28]. Associations between parameters were tested using Pearson’s correlation tests. All statistical analyses were performed using the statistical software packages SPSS 17.0 (SPSS Inc., Chicago, IL, USA) and R v.3.6.3 [29].

3. Results

3.1. Effect of the Presence and Type of Gene Overlapping on Gene Evolution

We analyzed the effect of gene overlapping on the rate of begomovirus evolution by estimating the total length of the tree (t) inferred for the OV and NOV regions of each gene (Figure 2). In OV regions, t ranged from 0.001 to 5.218, depending on the gene–virus combination, with mean value of 0.797 (median: 0.573). Variation in t for NOV regions ranged between 0.006 and 7.676, with mean value of 1.262 (median: 0.744). A GzLM analysis using type of region (OV and NOV) as a factor indicated that t was significantly smaller in OV than in NOV regions (Wald $\chi^2 = 10.74$; $p = 1 \times 10^{-3}$). In agreement with these results, in most overlapping instances, t was significantly smaller in OV than in NOV regions (93/125, $\chi^2 = 59.54$, $p < 1 \times 10^{-5}$) (Figure 2 and Appendix A). Hence, gene overlapping generally reduces evolutionary rates. However, viruses can generate different types of gene overlap, which arise by different mechanisms and that generally differ in the resulting frameshift [10] and the degree of selective independence of the genes involved [7]. Therefore, it could be hypothesized that evolutionary rates differ by type of gene overlap.

Thus, three types of gene overlap were defined following [1]: (1) internal overlapping, when one of the genes contains the complete sequence of the other; (2) 5′-terminal overlapping, when the OV region is in the 5′-terminal region of the gene; and (3) 3′-terminal overlapping, when the OV region is in the 3′-terminal region of the gene. Genes with

terminal overlapping showed significantly lower t values in OV than in NOV regions (Wald $\chi^2 \geq 4.88$; $p \leq 0.027$), with most instances fitting this general observation (42/54, $\chi^2 = 33.33$, $p < 1 \times 10^{-5}$; and 39/54, $\chi^2 = 21.33$, $p < 1 \times 10^{-5}$, for 5'- and 3'-terminal overlapping, respectively). In contrast, in genes with internal overlapping no significant differences between OV and NOV regions were observed (Wald $\chi^2 \leq 1.99$; $p \geq 0.212$), and instances with lower t in OV regions were not significantly more frequent (11/17, $\chi^2 = 2.94$, $p = 0.086$) (Figure 2 and Appendix A). We also analyzed differences in the magnitude of the effect of each type of overlapping in reducing the rate of virus evolution. For that, we calculated the OV/NOV ratio for t values of overlapping instances where this parameter was significantly smaller in OV regions. A GzLM indicated that the magnitude of the effect on t depended on the type of overlapping (Wald $\chi^2 = 3.71$; $p = 0.028$), with terminal ones showing similar effects ($p = 0.314$) and in both cases higher than internal overlapping ($p \leq 0.041$). Same conclusions were obtained when normalized t values were used.

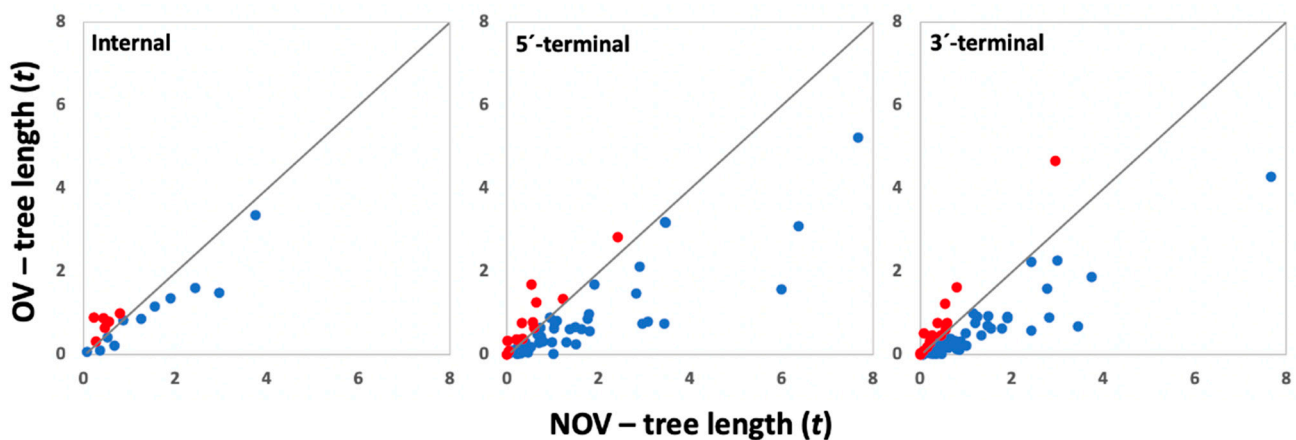


Figure 2. Overlapping and nonoverlapping tree lengths (t) in overlapping genes. Blue dots denote genes in which t is significantly higher in NOV than in OV regions. Red dots denote genes showing the opposite trend. Genes with different types of overlapping (internal, 5'-, and 3'-terminal) are presented in different panels.

Our dataset included mono- and bipartite begomoviruses, which differ in host-virus and virus-virus protein-protein interactions [30]. This may result in differential evolutionary constraints that may modulate how gene overlapping affects virus evolution. Thus, we analyzed whether gene overlapping influenced tree length depending on the begomovirus genome structure. GzLMs using this trait (mono- vs. bipartite) as a factor indicated that it had no effect on t differences between OV and NOV regions (Wald $\chi^2 = 2.24$; $p = 0.137$). In agreement, t was significantly higher in NOV than in OV regions when mono- and bipartite begomoviruses were analyzed separately (Wald $\chi^2 = 4.30$; $p = 0.038$ and Wald $\chi^2 = 8.67$; $p = 3 \times 10^{-3}$, respectively). In both groups of viruses, the same was observed when each type of terminal overlapping was analyzed separately (Wald $\chi^2 \geq 4.90$; $p \leq 0.027$), but not for internal overlapping (Wald $\chi^2 \leq 0.82$; $p \geq 0.366$). The proportion of instances with higher t in NOV than in OV regions was higher than expected by chance in terminal overlapping of both types of genome structures ($\chi^2 \geq 4.92$, $p \leq 0.026$), but not in internal overlapping ($\chi^2 \leq 0.98$, $p \geq 0.173$) (Appendix A).

In sum, these results indicate that the effect of gene overlapping on the rate of begomovirus evolution varies depending on its type; terminal overlapping generally reduces tree length, whereas no clear trend is observed in genes with internal overlapping. On the other hand, the type of genomic structure has little effect on the observed patterns.

3.2. Association between Selection Pressures and Gene Evolution

To further analyze how gene overlapping reduced the rate of evolution, we estimated selection pressures (d_N/d_S) and individual d_N and d_S values for the OV and NOV regions

of each gene (Figure 3 and Appendix A). Average d_N/d_S values were 0.35 ± 0.03 and 0.50 ± 0.04 for OV and NOV regions, respectively. A GzLM using type of region as factor indicated that negative selection pressures were significantly stronger in OV than in NOV regions (Wald $\chi^2 = 11.42$; $p < 1 \times 10^{-5}$), and we obtained similar results when each type of overlap was analyzed independently (Wald $\chi^2 \geq 8.18$; $p \leq 2 \times 10^{-4}$, Figure 3 and Appendix A). In agreement, most genes had significantly higher d_N/d_S in NOV than in OV fragments when all types of overlap were considered together (89/125, $\chi^2 = 44.94$, $p < 1 \times 10^{-5}$) and for the three of them independently (14/17, $\chi^2 = 14.24$, $p = 1.6 \times 10^{-4}$; 43/54, $\chi^2 = 37.93$, $p < 1 \times 10^{-5}$; 33/54, $\chi^2 = 5.33$, $p = 0.021$, for internal, 5'-, and 3'-overlapping, respectively) (Figure 3 and Appendix A).

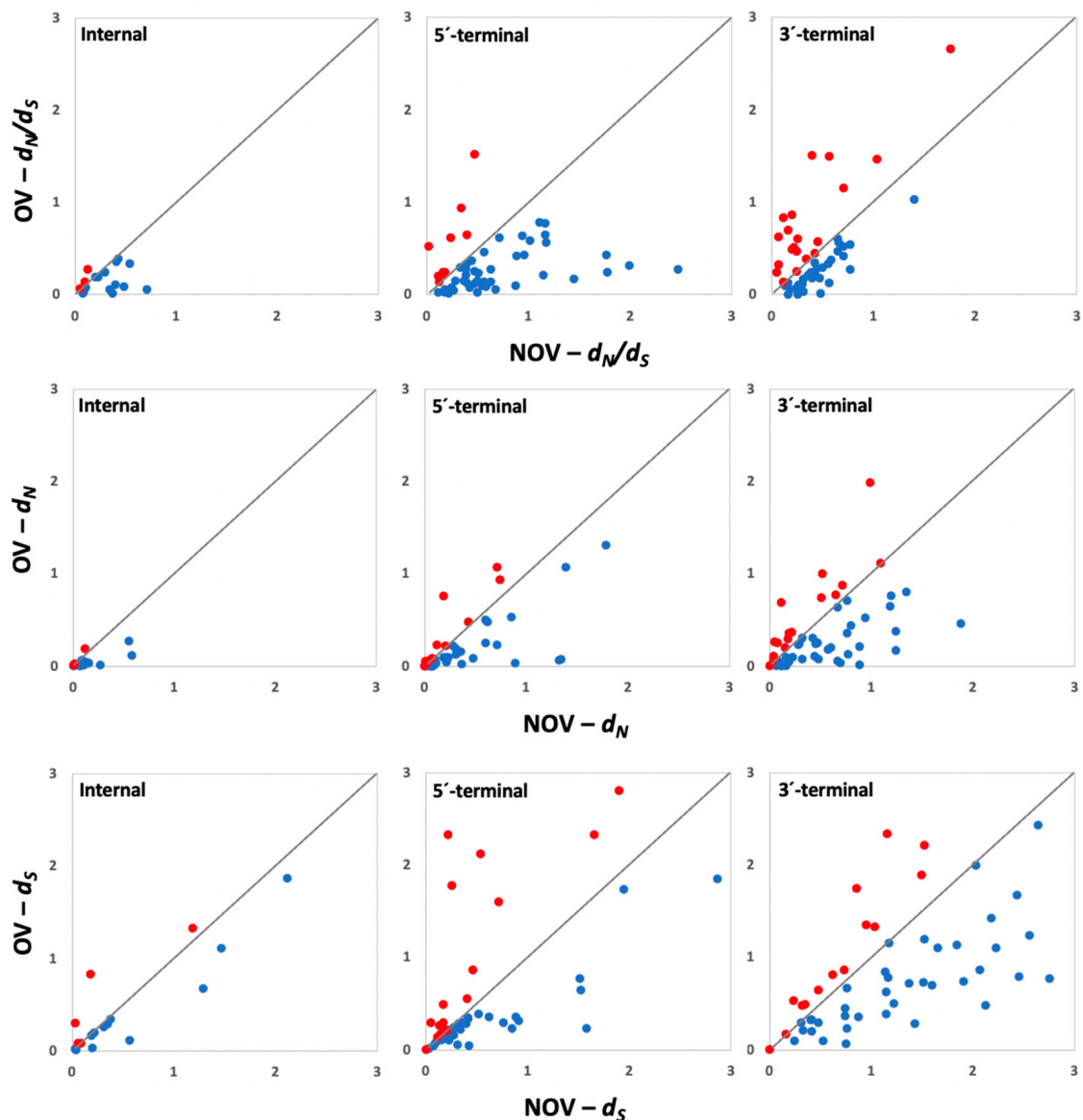


Figure 3. Overlapping and nonoverlapping d_N/d_S (upper line), d_N (middle line) and d_S (lower line) in overlapping genes. Blue dots denote instances in which parameters are significantly higher in NOV than in OV regions. Red dots denote genes showing the opposite trend. Genes with different types of overlapping (internal, 5'-, and 3'-terminal) are presented in different panels.

Similar analysis for d_N indicated significantly lower values in OV than in NOV regions when all genes were considered together (0.24 ± 0.03 and 0.39 ± 0.04 , respectively; Wald $\chi^2 = 12.10$; $p < 1 \times 10^{-5}$), and when each type of overlapping was analyzed separately

(Wald $\chi^2 \geq 9.21$; $p \leq 5 \times 10^{-3}$). Also, in most instances, d_N followed this trend (92/125, $\chi^2 = 55.70$, $p < 1 \times 10^{-5}$), with similar results for each type of overlap (13/17, $\chi^2 = 7.53$, $p = 6.1 \times 10^{-3}$; 42/54, $\chi^2 = 33.33$, $p < 1 \times 10^{-5}$; 37/54, $\chi^2 = 14.81$, $p = 1.2 \times 10^{-4}$, for internal, 5'-, and 3'-overlapping, respectively) (Figure 3). Finally, d_S was similar in NOV and in OV regions either considering all genes together (0.72 ± 0.08 and 0.83 ± 0.07 , respectively; Wald $\chi^2 = 2.16$; $p = 0.079$) or analyzing each type of overlap independently (Wald $\chi^2 \leq 3.18$; $p \geq 0.101$) (Figure 3). However, instances with d_S value higher in NOV than in OV regions were more frequent than expected by chance (86/125, $\chi^2 = 35.34$, $p < 1 \times 10^{-5}$), with similar results for each type of overlap (12/17, $\chi^2 = 5.76$, $p = 0.016$; 35/54, $\chi^2 = 9.48$, $p = 2.1 \times 10^{-3}$; 39/54, $\chi^2 = 21.33$, $p < 1 \times 10^{-5}$, for internal, 5'-, and 3'-overlapping, respectively) (Figure 3).

When mono- and bipartite begomoviruses were analyzed separately, d_N/d_S and d_N (Wald $\chi^2 \geq 7.70$; $p \leq 6 \times 10^{-3}$, Wald $\chi^2 \geq 4.18$; $p \leq 0.041$, respectively), but not d_S (Wald $\chi^2 \leq 3.00$; $p \geq 0.083$), were always higher in NOV than in OV regions for viruses with both genomic structures. When each type of overlapping was analyzed separately, similar results were obtained for d_N/d_S and d_N (Wald $\chi^2 \geq 4.11$; $p \leq 0.043$, Wald $\chi^2 \geq 6.61$; $p \leq 0.010$ and Wald $\chi^2 \geq 3.60$; $p \leq 0.050$, for internal, 5'-, and 3'-overlapping, respectively), and for d_S (Wald $\chi^2 \geq 0.20$; $p \leq 0.652$, Wald $\chi^2 \geq 1.20$; $p \leq 0.274$ and Wald $\chi^2 \geq 2.66$; $p \leq 0.103$, for internal, 5'-, and 3'-overlapping, respectively). As above, the proportion of overlapping instances with higher d_N/d_S , d_N and d_S in NOV than in OV regions was generally larger than those showing the opposite trend in viruses with both types of genome structure and in all types of gene overlapping ($\chi^2 \geq 3.63$, $p \leq 0.050$) (Appendix A).

Thus, overlapping genes are generally subjected to stronger purifying selection in OV than in NOV fragments, which seems to be associated with a greater constraint against non-synonymous changes regardless of the type of overlap and, to a lesser extent, with constraints to synonymous changes. Again, the type of genomic structure had no influence in the observed results.

3.3. Association between Proportion of Overlap and Gene Evolution

For RNA viruses, it has been shown that the lengths of the OV region relative to gene length are negatively correlated with these rates in a non-linear manner [1,31]. We analyzed whether this relationship held for begomoviruses by calculating the normalized tree length for the complete sequence of each gene and assessing the strength of association between t and the proportion of gene overlap (Figure 4). As the genome structure had no effect in previous analyses, we did not consider this trait here. On the other hand, we included normalized tree lengths for AC4 (100% overlap), which were not considered previously as in this gene no OV vs. NOV comparison was possible.

The proportion of gene overlap (%) differed among types of overlap (Wald $\chi^2 = 8.74$; $p < 1 \times 10^{-4}$): it was lower in genes with internal (35.67 ± 0.97) than terminal overlapping (60.59 ± 3.65 and 59.90 ± 3.18 for 5'- and 3'-terminal overlapping, respectively). Hence, we analyzed the association between per cent of gene overlap and t in the complete sequence of each gene for all genes together and for each type of overlap separately. We performed bivariate analysis considering linear and nonlinear regressions. When a significant association was found, it was best explained by a negative logarithmic relationship between the length of overlap and t (Figure 4). Bivariate analysis revealed a significant negative logarithmic association between these two variables when all instances were considered together ($r = -0.33$; $p < 1 \times 10^{-4}$; Figure 4), with similar results when excluding values for AC4 ($r = -0.32$; $p < 1 \times 10^{-4}$). We also found a significant negative logarithmic association in both types of terminal overlap ($r = -0.37$, $p = 9 \times 10^{-3}$ and $r = -0.31$, $p = 0.027$; for 5'-, and 3'-overlapping, respectively), but not for internal ones with ($r = -0.25$, $p = 0.191$; Figure 4) and without ($r = 0.23$, $p = 0.383$) AC4 values. Comparable results were obtained using only those genes for which t values were higher in NOV than in OV regions.

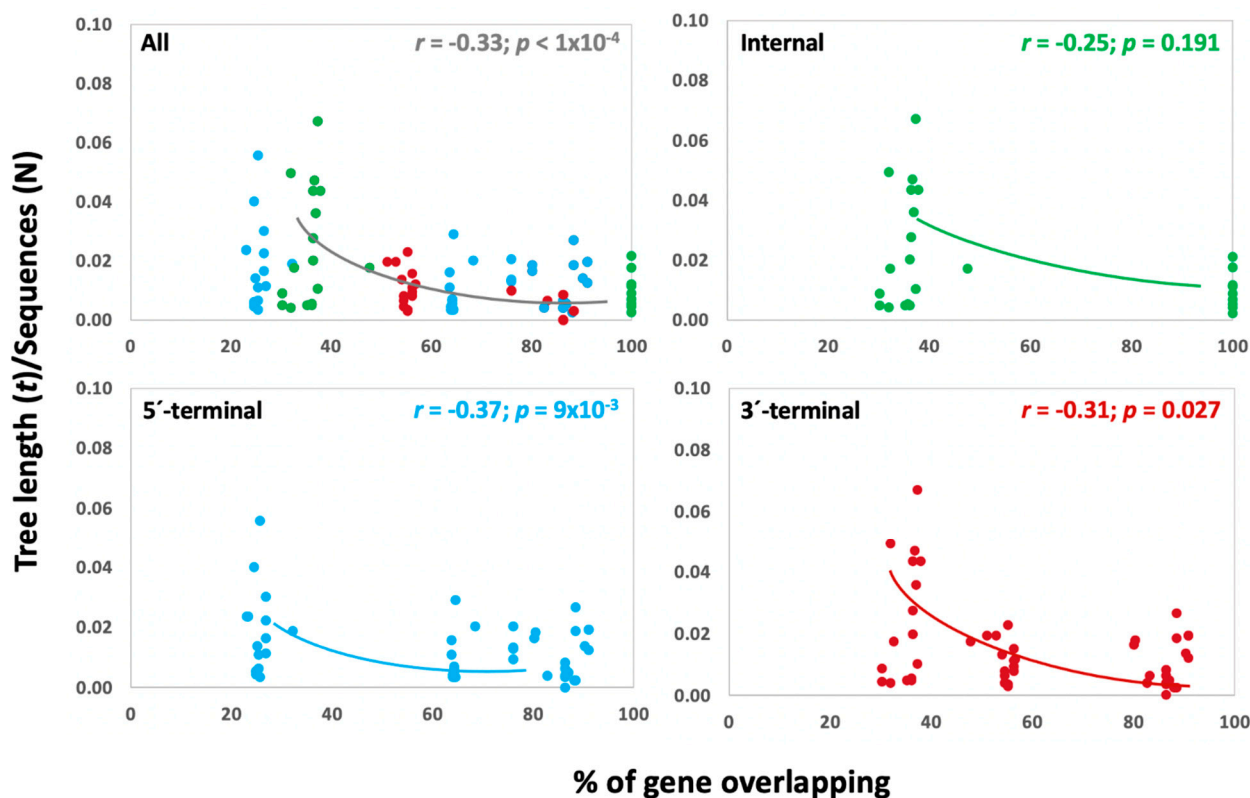


Figure 4. Correlation between tree length (t /number of sequences) and the proportion of gene overlap (length of overlap/total gene length) for all types of overlap considered together (**upper left**), internal overlapping (**upper right**), 5'-terminal overlapping (**lower left**) and 3'-terminal overlapping (**lower right**).

4. Discussion

Several non-mutually exclusive theories have been proposed to explain the abundance of gene overlapping in viruses: (i) it has a role in gene regulation by providing an inherent mechanism for coordinated expression [7]; (ii) it is an effective mechanism for generating novel genes while keeping genome size minimized, by introducing a new reading frame on top of an existing one [32,33]; or (iii) as mutations in these regions affect more than one gene, gene overlapping amplifies the deleterious effect of mutations, thus quickly eliminating such mutations from the viral population, particularly in RNA viruses which have higher mutation rates [7,34,35]. Although there is general agreement on the role of gene overlapping in maintaining genomic compression [10,31,36,37], its effect on virus evolutionary rates remains more elusive [1,2]. This is particularly so for DNA viruses that despite having in general lower mutation rates than RNA viruses have in some cases larger proportion of their genes with at least one overlapping instance [2]. Here, we analyzed whether in the largest genus of plant DNA viruses, whose genome is enriched in gene overlapping instances, this feature modulates the rate of gene evolution.

Our comparative genomic analyses in species of the genus *Begomovirus* indicate that tree length (as a proxy of the rate of evolution) was generally smaller in OV than in NOV regions, with most overlapping instances following this rule. This agrees with the predictions of mathematical models [34,38,39]. Interestingly, these models also predict that the reduction in evolutionary rate is the consequence of correlations at overlapping sites, which are stronger in positions where a mutation would result in a nonsynonymous change in both overlapping genes than in positions where mutations are synonymous in one gene and nonsynonymous in the other [7,34,38]. This may explain why our results indicate that the reduction in the genetic diversity of OV regions is associated with decreased d_N , but not d_S although in most instances OV regions had lower values of both parameters:

gene overlapping would influence both synonymous and nonsynonymous substitution, but this effect would be stronger in nonsynonymous ones. There was significant negative (logarithmic) correlation between the length of overlap and the genetic diversity of each gene; that is, the longer the OV region, the lower the evolutionary rate. This agrees with theoretical models, which predict that evolutionary rate is expected to decline nonlinearly with increasing overlap [7]. This negative logarithmic association also indicates that an increased proportion of gene overlapping reduces begomovirus evolutionary rates up to a threshold, beyond which larger overlapping has no effect on tree length. Thus, long overlapping regions cannot be fully explained by their effect on evolutionary rates alone, and other selection pressures, such as genome compression or coordinated gene expression are likely to play a role.

Altogether, our results provide compelling evidence supporting the role of gene overlapping in reducing the rate of *Begomovirus* evolution. This observation is in accordance with previous reports for a variety of RNA viruses [1,40–42]. In most of these cases, the reduction of the rate of virus evolution associated with gene overlapping has been attributed to the need of these viruses to buffer excessive mutational load due to high mutation rates. To date, however, estimates of mutation rates in DNA viruses suggest that these are lower than for RNA viruses [11]. Two lines of evidence suggest that this might not be the case for begomoviruses. First, rough estimates of mutation frequency in TYLCCNV showed values around 1×10^{-4} [19], which is comparable to the variation reported for plant RNA viruses and higher than for other ssDNA viruses [11,43]. Second, it has been shown that some of the DNA polymerases involved in begomovirus replication are error-prone in conditions equivalent to those in which they amplify the viral genome [44,45]. Hence, begomoviruses could have evolved overlapping regions as a safety mechanism to control high mutation rates.

Evolutionary constraints imposed by gene overlapping are a double-edged sword. They restrict the fixation of deleterious mutations; but at the same time, they leave little room to increase virus fitness, as beneficial mutations in one gene are often deleterious in the other and are therefore purged [1,4]. Viruses are faced with the need to reconcile these two facets such that they limit the fixation of unfit mutations but allow generation of beneficial genetic diversity. To do so, it has been shown that viruses may use a “segregated” organization in which overlapped regions harbor functional domains of one gene or the other, but never both [4]. Thus, gene overlapping imposes a certain degree of evolutionary constraint, as mutations affect more than one gene at the same time. However, this is not as strong as if both genes would harbor functional domains in the overlapping region, or as relaxed as if both genes would not overlap. This strategy results in higher fitness peaks than in the absence of gene overlapping [4]. Interestingly, some evidence suggests that begomoviruses may use a similar strategy. For instance, AV1 functional domains involved in DNA shuttle into the nucleus or in vector transmission are located at the N-terminal region of the protein, overlapping with AV2 [46]; whereas hydrophobic domains involved in the silencing suppression activity of AV2 locate at the NOV region of this protein [47]. Similarly, in AC2 the domain responsible for repressing AV1 expression is in the NOV region of this gene [48], whereas the OV region of AC3 is rich in functional domains [47].

Despite the general trend toward a reduction of genetic diversity in OV compared with NOV regions, when each type of overlapping was analyzed separately, this effect remained significant only in instances with 5′- and 3′-terminal overlap, whereas nearly one-third of the instances with internal overlap showed the opposite trend. Different types of overlapping vary in the preponderance of the associated frameshifting [10]. However, in begomoviruses all overlapping instances have +1/−1 frameshift, which are identical in the extent to which they allow selective independence of the overlapping genes [7]. Alternatively, in our dataset we included mono- and bipartite begomoviruses, for which different functions have been attributed to the C4/AC4 proteins [30]. Different selective pressures on C4/AC4 depending on its function may impose different constraints on its evolution, modulating the buffering effects of gene overlapping on the accumulation of

mutations on AC1. We do not favor this hypothesis as our results indicate that genomic structure has little effect on the role of gene overlapping as modulator of begomovirus evolution. Another possible explanation for the observed differences is that, as we restricted our analyses to a single virus genus and each type of overlap occurs in the same genes across species, differences between terminal and internal overlapping reflect particular characteristics of the genes involved. Indeed, internal overlapping instances involved the same two genes (AC1 and AC4) in all species. If, for instance, the AC1 gene is dominating the evolution of AC4, as has been shown for younger overlapping genes generated by overprinting over older ones [33,49], the resulting internal overlapping would have less effect in the evolution of AC1, in accordance with our results. In addition, note that AC1 is involved in virus replication, which is a key component of virus fitness, thus this gene is more likely to drive AC4 evolution rather than the other way around. In support of this hypothesis, it has been shown that AC1 is under strong negative selection, whereas AC4 is under positive selection [50]. Finally, at odds with the examples mentioned above, functional conserved domains are not segregated in AC1/AC4 [47,51], which would also support that the observed differences respond to gene-specific features.

An additional source of gene-specific heterogeneity in our dataset that could explain the differential effect of internal and terminal overlapping in begomovirus evolution is the presence of recombination. Large fragments of AC1 (including the region overlapping with AC4) are recombination hotspots, whereas AC2/AC3 and big portions of AV1 and AV2 are coldspots [26,52]. It has been hypothesized that recombination allows removing deleterious mutations with high efficiency, as reviewed by [53]. Hence, the limited effect of AC1/AC4 internal overlapping in virus evolutionary rates could be explained by a higher frequency of recombination in AC1, which in NOV regions would have similar consequences than gene overlapping. Although we cannot completely discard such a role of recombination, at least in our dataset several observations argue against it. First, the percentage of instances with over 10% of recombinant sequences was evenly distributed across types of overlapping (31–35%, Appendix A). If recombination in AC1/AC4 were to explain our results, we would have expected more frequent recombination in internal than in terminal overlapping. Rather, virus species identity seemed to explain most of the variation in recombination frequency, with three species (*Bhendi yellow vein mosaic virus*, *Chilli leaf curl virus* and *Okra enation leaf curl virus*) accounting for two thirds of the overlapping instances with excessive recombination. Second, when these instances (41/125) were removed from the analyses, we still observed higher *t* values in NOV than in OV regions in terminal (30/37 and 27/36 instances, for 5'-, and 3'-terminal overlapping, respectively), but not in internal (5/10 instances), overlapping. Hence, our conclusions hold regardless of the presence of extensive recombination.

Some cautionary comments on our results are called for, however. First, the number of instances among types of overlapping is not fully balanced, with lower numbers for internal than for both types of terminal overlap. Hence, the lack of a significant effect of OV regions in internal overlap could be due to reduced sample size. Second, because we restricted our analyses to a single virus genus, overlapping instances occur in the same genes across species, which may reduce the range of overlapping lengths included in the regression analyses with the subsequent reduction of statistical power. However, the range of terminal overlapping lengths included was enough to detect a significant correlation between percent of gene overlapping and genetic diversity. This range was much smaller for internal overlapping, which again may explain the lack of association between the two analyzed traits. Finally, we could include only 18 out of the 420 begomovirus species, as these were the only ones that fulfilled the criteria to be included in our analyses. Despite the small sample size, our results indicate strongly significant effects, which support the relevant role of gene overlapping in begomovirus evolution.

In sum, this work provides novel evidence of the selective constraints imposed by gene overlapping on the pace of begomovirus evolution. Whether this effect is general for DNA viruses would be an interesting avenue of future research.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/microorganisms10020366/s1>, File S1: Sequence alignments used in this work.

Author Contributions: Conceptualization, I.P.; methodology, I.P.; formal analysis, I.M.-H.; data curation, I.M.-H.; writing—original draft preparation, I.P.; writing—review and editing, I.P.; funding acquisition, I.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Plan Nacional I + D + i, Ministerio de Economía y Competitividad (Agencia Estatal de Investigación), Spain [PID2019-109579RB-I00] to IP.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available as Supplementary Material and as Appendix A.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Table A1. Statistical parameters of traits associated with the rate of evolutionary change in overlapping genes.

Species ¹	N ²	Overlapping ³		Gene Length ⁴	t ⁵		d _N /d _S	
		ORF	Length		NOV	OV	NOV	OV
Internal (11/17)								
<i>African cassava mosaic virus (B)</i>	27 (4%)	AC1-AC4	423	1077	0.53	0.41	0.39	0.11
<i>Alternanthera yellow vein virus (M)</i>	11 (8%)	AC1-AC4	291	1086	0.47	0.64	0.54	0.33
<i>Bean golden mosaic virus (B)</i>	121 (0%)	AC1-AC4	258	1086	0.44	0.86	0.34	0.06
<i>Bhendi yellow vein mosaic virus* (M)</i>	39 (98%)	AC1-AC4	294	1092	1.25	0.85	0.48	0.09
<i>Bhendi yellow vein mosaic virus (M)</i>	32 (99%)	AC1-AC4	303	1092	1.90	1.35	0.11	0.08
<i>Chilli leaf curl virus (M)</i>	16 (28%)	AC1-AC4	300	1086	0.86	0.83	0.22	0.20
<i>Cotton leaf curl Gezira virus (M)</i>	21 (9%)	AC1-AC4	294	1089	0.22	0.89	0.37	0.02
<i>Cotton leaf curl Multan virus (M)</i>	50 (2%)	AC1-AC4	303	1092	0.67	0.21	0.07	0.02
<i>East African cassava mosaic Kenya virus (B)</i>	50 (0%)	AC1-AC4	297	1065	0.27	0.31	0.05	0.07
<i>East African cassava mosaic Malawi virus (B)</i>	13 (0%)	AC1-AC4	234	1080	0.07	0.06	0.20	0.19
<i>East African cassava mosaic virus (B)</i>	153 (2%)	AC1-AC4	234	1080	1.55	1.15	0.41	0.36
<i>East African cassava mosaic Zanzibar virus (B)</i>	14 (29%)	AC1-AC4	258	1080	0.35	0.09	0.08	0.03
<i>Okra enation leaf curl virus (M)</i>	60 (92%)	AC1-AC4	308	1089	2.96	1.48	0.29	0.24
<i>South African cassava mosaic virus (B)</i>	125 (0%)	AC1-AC4	297	1080	0.55	0.79	0.09	0.14
<i>Sweet potato leaf curl virus (M)</i>	17 (41%)	AC1-AC4	258	1095	0.79	0.99	0.12	0.28
<i>Tomato leaf curl New Delhi virus (B)</i>	97 (3%)	AC1-AC4	303	1086	3.75	3.34	0.43	0.39
<i>Tomato yellow leaf curl virus (M)</i>	397 (68%)	AC1-AC4	294	1074	2.43	1.60	0.70	0.05
5'-terminal (42/54)								
<i>African cassava mosaic virus (B)</i>	32 (3%)	AC1-AC2	93	408	0.21	0.37	0.11	0.20
	31 (3%)	AC2-AC3	260	405	0.31	0.03	0.50	0.03
	26 (0%)	AV1-AV2	193	777	0.39	0.26	0.37	0.15
<i>Alternanthera yellow vein virus (M)</i>	11 (18%)	AC1-AC2	98	406	0.22	0.01	0.22	0.02
	13 (0%)	AC2-AC3	260	405	0.71	0.27	0.11	0.03
	10 (60%)	AV1-AV2	189	771	0.33	0.76	1.15	0.21
<i>Bean golden mosaic virus (B)</i>	158 (0%)	AC1-AC2	89	390	1.30	0.29	0.40	0.13
	158 (0%)	AC2-AC3	254	399	1.50	0.25	0.26	0.05
<i>Bhendi yellow vein mosaic virus* (M)</i>	51 (71%)	AC1-AC2	104	453	1.91	1.68	0.33	0.29
	51 (20%)	AC2-AC3	308	405	3.44	0.74	0.28	0.14
	50 (40%)	AV1-AV2	206	771	1.76	0.85	0.63	0.14
<i>Bhendi yellow vein mosaic virus (M)</i>	57 (54%)	AC1-AC2	104	453	2.42	2.82	1.01	0.58
	57 (2%)	AC2-AC3	308	405	1.80	0.56	1.18	0.56
	56 (16%)	AV1-AV2	206	771	1.63	0.61	0.38	0.21
<i>Chilli leaf curl virus (M)</i>	18 (33%)	AC1-AC2	98	405	0.99	0.30	0.17	0.02
	23 (17%)	AC2-AC3	260	405	3.08	0.79	0.47	0.25
	22 (18%)	AV1-AV2	197	771	1.22	1.33	0.18	0.24
<i>Cotton leaf curl Gezira virus (M)</i>	32 (6%)	AC1-AC2	101	405	0.01	0.33	0.02	0.52
	32 (0%)	AC2-AC3	257	402	0.47	0.04	0.42	0.07
	31 (0%)	AV1-AV2	209	777	0.35	0.37	0.16	0.24

Table A1. Cont.

Species ¹	N ²	Overlapping ³		Gene Length ⁴	<i>t</i> ⁵		<i>d_N/d_S</i>	
		ORF	Length		NOV	OV	NOV	OV
<i>Cotton leaf curl Multan virus (M)</i>	58 (2%)	AC1-AC2	104	453	0.59	0.65	0.12	0.14
	59 (2%)	AC2-AC3	308	405	1.08	0.80	1.11	0.78
	59 (15%)	AV1-AV2	206	771	1.04	0.75	0.44	0.37
<i>East African cassava mosaic Kenya virus (B)</i>	71 (0%)	AC1-AC2	77	408	0.40	0.29	0.25	0.08
	71 (0%)	AC2-AC3	260	405	0.34	0.22	0.18	0.06
	64 (0%)	AV1-AV2	197	774	0.23	0.18	0.55	0.10
<i>East African cassava mosaic Malawi virus (B)</i>	9 (0%)	AC1-AC2	92	408	0.00	0.00	-	-
	10 (0%)	AC2-AC3	260	405	0.06	0.03	0.57	0.14
	12 (0%)	AV1-AV2	191	777	0.05	0.08	0.17	0.22
<i>East African cassava mosaic virus (B)</i>	166 (2%)	AC1-AC2	92	405	1.79	0.97	2.47	0.27
	162 (0%)	AC2-AC3	260	405	1.38	0.60	0.67	0.05
	105 (0%)	AV1-AV2	197	775	0.79	0.31	0.58	0.09
<i>East African cassava mosaic Zanzibar virus (B)</i>	15 (27%)	AC1-AC2	92	408	0.23	0.01	0.50	0.15
	15 (0%)	AC2-AC3	260	405	0.11	0.10	1.76	0.42
	15 (0%)	AV1-AV2	197	775	0.20	0.07	0.63	0.27
<i>Okra enation leaf curl virus (M)</i>	67 (97%)	AC1-AC2	104	453	3.45	3.18	0.51	0.23
	68 (32%)	AC2-AC3	308	405	1.04	0.63	1.44	0.17
	41 (58%)	AV1-AV2	188	771	0.64	1.26	0.24	0.62
<i>Pepper golden mosaic virus (B)</i>	54 (57%)	AC1-AC2	59	390	1.49	0.66	0.47	0.12
	54 (0%)	AC2-AC3	254	399	0.94	0.89	0.34	0.94
<i>Pepper huasteco yellow vein virus (B)</i>	19 (5%)	AC1-AC2	80	417	1.01	0.01	0.38	0.28
	45 (0%)	AC2-AC3	254	399	0.74	0.42	1.99	0.32
<i>South African cassava mosaic virus (B)</i>	131 (0%)	AC1-AC2	92	408	0.56	0.77	0.47	1.52
	130 (0%)	AC2-AC3	260	405	0.67	0.48	0.40	0.64
	126 (4%)	AV1-AV2	191	777	0.73	0.64	1.17	0.65
<i>Sweet potato leaf curl virus (M)</i>	18 (39%)	AC1-AC2	92	452	0.54	1.68	0.39	0.33
	14 (7%)	AC2-AC3	278	435	2.96	0.73	1.78	0.25
	18 (9%)	AV1-AV2	176	765	0.52	0.17	0.87	0.10
<i>Tomato leaf curl New Delhi virus (B)</i>	88 (4%)	AC1-AC2	98	420	2.83	1.47	0.71	0.62
	113 (1%)	AC2-AC3	281	411	5.99	1.56	0.57	0.46
	115 (0%)	AV1-AV2	179	771	2.90	2.12	1.16	0.77
<i>Tomato yellow leaf curl virus (M)</i>	588 (9%)	AC1-AC2	92	408	7.68	5.22	0.96	0.43
	521 (10%)	AC2-AC3	260	405	6.37	3.08	0.95	0.64
	593 (10%)	AV1-AV2	191	777	3.47	3.16	0.89	0.42
3'-terminal (39/54)								
<i>African cassava mosaic virus (B)</i>	27 (3%)	AC1-AC2	93	1077	0.53	0.36	0.50	0.29
	32 (3%)	AC2-AC3	260	408	0.21	0.03	0.44	0.20
	33 (0%)	AV1-AV2	193	342	0.54	0.24	0.47	0.18
<i>Alternanthera yellow vein virus (M)</i>	11 (18%)	AC1-AC2	98	1086	0.47	0.01	0.48	0.02
	11 (0%)	AC2-AC3	260	406	0.22	0.27	0.45	0.57
	12 (60%)	AV1-AV2	189	348	1.21	0.76	0.57	0.13
<i>Bean golden mosaic virus (B)</i>	121 (0%)	AC1-AC2	89	1086	0.44	0.10	0.56	0.34
	158 (0%)	AC2-AC3	254	390	0.30	0.28	0.65	0.61
<i>Bhendi yellow vein mosaic virus* (M)</i>	39 (92%)	AC1-AC2	104	1092	1.25	0.91	0.58	0.38
	51 (20%)	AC2-AC3	308	453	1.91	0.91	0.27	0.03
	50 (40%)	AV1-AV2	206	366	1.34	0.46	0.13	0.10
<i>Bhendi yellow vein mosaic virus (M)</i>	23 (98%)	AC1-AC2	104	1092	1.90	0.88	0.16	0.70
	57 (2%)	AC2-AC3	308	453	2.42	0.57	0.21	0.51
	55 (16%)	AV1-AV2	206	366	0.89	0.32	0.78	0.54
<i>Chilli leaf curl virus (M)</i>	16 (38%)	AC1-AC2	98	1086	0.86	0.10	0.30	0.12
	18 (17%)	AC2-AC3	260	405	0.99	0.50	0.66	0.57
	16 (19%)	AV1-AV2	197	357	0.26	0.45	1.76	2.67
<i>Cotton leaf curl Gezira virus (M)</i>	21 (9%)	AC1-AC2	101	1089	0.22	0.33	0.07	0.33
	32 (0%)	AC2-AC3	257	405	0.01	0.03	0.24	0.47
	29 (0%)	AV1-AV2	209	369	0.08	0.50	0.07	0.63
<i>Cotton leaf curl Multan virus (M)</i>	50 (2%)	AC1-AC2	104	1092	0.67	0.38	0.27	0.03
	58 (2%)	AC2-AC3	308	453	0.59	0.75	0.11	0.14
	57 (16%)	AV1-AV2	206	366	0.45	0.46	0.24	0.26
<i>East African cassava mosaic Kenya virus (B)</i>	50 (0%)	AC1-AC2	77	1065	0.27	0.01	0.38	0.25
	71 (0%)	AC2-AC3	260	408	0.40	0.25	0.25	0.61
	56 (0%)	AV1-AV2	197	357	0.16	0.18	0.42	0.45
<i>East African cassava mosaic Malawi virus (B)</i>	13 (0%)	AC1-AC2	92	1080	0.07	0.00	0.16	0.00
	9 (0%)	AC2-AC3	260	408	0.00	0.00	-	-
	12 (0%)	AV1-AV2	191	351	0.06	0.08	0.34	0.39

Table A1. Cont.

Species ¹	N ²	Overlapping ³		Gene Length ⁴	t ⁵		d _N /d _S	
		ORF	Length		NOV	OV	NOV	OV
<i>East African cassava mosaic virus (B)</i>	153 (3%)	AC1-AC2	92	1080	1.55	0.62	0.43	0.25
	166 (0%)	AC2-AC3	260	405	1.79	0.62	0.43	0.35
	136 (0%)	AV1-AV2	197	357	0.75	0.32	0.36	0.22
<i>East African cassava mosaic Zanzibar virus (B)</i>	14 (29%)	AC1-AC2	92	1080	0.35	0.01	0.32	0.17
	15 (0%)	AC2-AC3	260	408	0.23	0.10	0.77	0.28
	13 (0%)	AV1-AV2	197	357	0.15	0.07	0.65	0.47
<i>Okra enation leaf curl virus (M)</i>	60 (92%)	AC1-AC2	104	1089	2.96	4.66	0.05	0.25
	67 (33%)	AC2-AC3	308	453	3.45	0.68	0.32	0.04
	45 (53%)	AV1-AV2	188	348	0.38	0.75	0.21	0.50
<i>Pepper golden mosaic virus (B)</i>	54 (57%)	AC1-AC2	59	1050	1.47	0.68	0.25	0.01
	54 (0%)	AC2-AC3	254	390	1.49	0.92	0.44	0.29
<i>Pepper huasteco yellow vein virus (B)</i>	44 (2%)	AC1-AC2	80	1050	0.81	0.12	0.26	0.11
	19 (0%)	AC2-AC3	254	417	1.01	0.21	1.40	1.03
<i>South African cassava mosaic virus (B)</i>	125 (0%)	AC1-AC2	92	1080	0.55	0.60	1.04	1.47
	131 (0%)	AC2-AC3	260	408	0.56	0.42	0.70	0.43
	128 (4%)	AV1-AV2	191	351	1.16	0.99	0.24	0.08
<i>Sweet potato leaf curl virus (M)</i>	17 (41%)	AC1-AC2	92	1095	0.79	1.61	0.57	1.50
	18 (6%)	AC2-AC3	278	452	0.54	1.21	0.70	1.16
	18 (9%)	AV1-AV2	176	345	0.62	0.16	0.18	0.06
<i>Tomato leaf curl New Delhi virus (B)</i>	97 (3%)	AC1-AC2	98	1086	3.75	1.86	0.16	0.13
	88 (1%)	AC2-AC3	281	420	2.83	0.88	0.12	0.84
	105 (0%)	AV1-AV2	179	339	2.78	1.58	0.71	0.52
<i>Tomato yellow leaf curl virus (M)</i>	397 (8%)	AC1-AC2	92	1074	2.43	2.23	0.40	1.52
	588 (9%)	AC2-AC3	260	408	7.67	4.27	0.20	0.87
	623 (9%)	AV1-AV2	191	351	3.01	2.27	0.40	0.18

¹ Number of genes with significant differences in tree length between OV and NOV regions over total number. Deviation from randomness was tested by Fisher's exact test ($p < 0.05$). M: Monopartite; B: Bipartite. Asterisks indicate sequences formerly classified as *Bendhi yellow vein India virus*; ² Number of sequences used. Percentage of recombinant sequences is shown in parentheses; ³ Name of the genes involved in the overlapping instance, and length of the OV region; ⁴ Length of the largest gene (internal overlapping) of the gene with 5'-terminal overlapping (5'-terminal) and of the gene with 3'-terminal overlapping (3'-terminal); ⁵ Tree length of inferred phylogenies for OV and NOV regions of each overlapping instance.

References

1. Simon-Loriere, E.; Holmes, E.C.; Pagán, I. The effect of gene overlapping on the rate of RNA virus evolution. *Mol. Biol. Evol.* **2013**, *30*, 1916–1928. [[CrossRef](#)] [[PubMed](#)]
2. Schlub, T.E.; Holmes, E.C. Properties and abundance of overlapping genes in viruses. *Virus Evol.* **2020**, *6*, veaa009. [[CrossRef](#)] [[PubMed](#)]
3. Gogarten, J.P.; Townsend, J.P. Horizontal Gene Transfer, Genome Innovation and Evolution. *Nat. Rev. Microbiol.* **2005**, *3*, 679–687. [[CrossRef](#)] [[PubMed](#)]
4. Fernandes, J.D.; Faust, T.B.; Strauli, N.B.; Smith, C.; Crosby, D.C.; Nakamura, R.L.; Hernandez, R.D.; Frankel, A.D. Functional Segregation of Overlapping Genes in HIV. *Cell* **2016**, *167*, 1762–1773. [[CrossRef](#)] [[PubMed](#)]
5. Nelson, C.W.; Ardern, Z.; Goldberg, T.L.; Meng, C.; Kuo, C.-H.; Ludwig, C.; Kolokotronis, S.-O.; Wei, X. Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *eLife* **2020**, *9*, e59633. [[CrossRef](#)] [[PubMed](#)]
6. Wright, B.W.; Ruan, J.; Molloy, M.P.; Jaschke, P.R. Genome modularization reveals overlapped gene topology is necessary for efficient viral reproduction. *ACS Synth. Biol.* **2020**, *9*, 3079–3090. [[CrossRef](#)]
7. Krakauer, D.C. Stability and evolution of overlapping genes. *Evolution* **2000**, *54*, 731–739. [[CrossRef](#)]
8. Belshaw, R.; Gardner, A.; Rambaut, A.; Pybus, O.G. Pacing a small cage: Mutation and RNA viruses. *Trends Ecol. Evol.* **2008**, *23*, 188–193. [[CrossRef](#)]
9. Brandes, N.; Linial, M. Gene overlapping and size constraints in the viral world. *Biol. Direct* **2016**, *11*, 26. [[CrossRef](#)]
10. Belshaw, R.; Pybus, O.G.; Rambaut, A. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* **2007**, *17*, 1496–1504. [[CrossRef](#)]
11. Sanjuán, R.; Domingo-Calap, P. Mechanisms of viral mutation. *Cell Mol. Life Sci.* **2016**, *73*, 4433–4448. [[CrossRef](#)]
12. Duffy, S.; Shackelton, L.A.; Holmes, E.C. Rates of evolutionary change in viruses: Patterns and determinants. *Nat. Rev. Genet.* **2008**, *9*, 267–276. [[CrossRef](#)] [[PubMed](#)]
13. Cui, J.; Schlub, T.E.; Holmes, E.C. An Allometric Relationship between the Genome Length and Virion Volume of Viruses. *J. Virol.* **2014**, *88*, 6403–6410. [[CrossRef](#)] [[PubMed](#)]
14. Wu, M.; Wei, H.; Tan, H.; Pan, S.; Liu, Q.; Bejarano, E.R.; Lozano-Durán, R. Plant DNA polymerases α and δ mediate replication of geminiviruses. *Nat. Commun.* **2021**, *12*, 2780. [[CrossRef](#)] [[PubMed](#)]

15. Fiallo-Olivé, E.; Lett, J.-M.; Martin, D.P.; Roumagnac, P.; Varsani, A.; Zerbini, F.M.; Navas-Castillo, J. ICTV Report Consortium. ICTV Virus Taxonomy Profile: *Geminiviridae* 2021. *J. Gen. Virol.* **2021**, *102*, 001696. [[CrossRef](#)] [[PubMed](#)]
16. Hanley-Bowdoin, L.; Bejarano, E.R.; Robertson, D.; Mansoor, S. Geminiviruses: Masters at redirecting and reprogramming plant processes. *Nat. Rev. Microbiol.* **2013**, *11*, 777–788. [[CrossRef](#)]
17. Zerbini, F.M.; Briddon, R.W.; Idris, A.; Martin, D.P.; Moriones, E.; Navas-Castillo, J.; Rivera-Bustamante, R.; Roumagnac, P.; Varsani, A.; ICTV Report Consortium. ICTV Virus Taxonomy Profile: *Geminiviridae*. *J. Gen. Virol.* **2017**, *98*, 131–133. [[CrossRef](#)]
18. Pagán, I.; García-Arenal, F. Population Genomics of Plant Viruses. In *Population Genomics: Microorganisms*; Polz, M.F., Rajora, O.P., Eds.; Springer International Publishing: New York, NY, USA, 2018.
19. Ge, L.; Zhang, J.; Zhou, X.; Li, H. Genetic Structure and Population Variability of Tomato Yellow Leaf Curl China Virus. *J. Virol.* **2007**, *81*, 5902–5907. [[CrossRef](#)]
20. Rodelo-Urrego, M.; García-Arenal, F.; Pagán, I. The effect of ecosystem biodiversity on virus genetic diversity depends on virus species: A study of chiltepin-infecting begomoviruses in Mexico. *Virus Evol.* **2015**, *1*, vev004. [[CrossRef](#)]
21. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)]
22. Larsson, A. AliView: A fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* **2014**, *30*, 3276–3278. [[CrossRef](#)] [[PubMed](#)]
23. Kosakovsky Pond, S.L.; Frost, S.D. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **2005**, *21*, 2531–2533. [[CrossRef](#)] [[PubMed](#)]
24. Kosakovsky Pond, S.L.; Frost, S.D.; Grossman, Z.; Gravenor, M.B.; Richman, D.D.; Leigh Brown, A.J. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comp. Biol.* **2006**, *2*, e62. [[CrossRef](#)] [[PubMed](#)]
25. Martin, D.P.; Varsani, A.; Roumagnac, P.; Botha, G.; Maslamoney, S.; Schwab, T.; Kelz, Z.; Kumar, V.; Murrell, B. RDP5: A computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* **2020**, *7*, veaa087. [[CrossRef](#)] [[PubMed](#)]
26. Lefeuvre, P.; Lett, J.-M.; Reynaud, B.; Martin, D.P. Avoidance of protein fold disruption in natural virus recombinants. *PLoS Pathog* **2007**, *3*, e181. [[CrossRef](#)] [[PubMed](#)]
27. Belgorodski, N.; Greiner, M.; Tolksdorf, K.; Schueller, K.; riskDistributions: Fitting Distributions to Given Data or Known Quantiles. R Package v.2.1. Available online: <https://CRAN.R-project.org/package=riskDistributions> (accessed on 16 December 2021).
28. Sokal, R.R.; Rohlf, F.J. *Biometry: The Principles and Practices of Statistics in Biological Research*; W.H. Freeman & Company: New York, NY, USA, 1995.
29. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021; Available online: <https://www.R-project.org/> (accessed on 16 December 2021).
30. Luna, A.P.; Lozano-Durán, R. Geminivirus-Encoded Proteins: Not All Positional Homologs Are Made Equal. *Front. Microbiol.* **2020**, *11*, 878. [[CrossRef](#)]
31. Chirico, N.; Vianelli, A.; Belshaw, R. Why genes overlap in viruses. *Proc. Biol. Sci.* **2010**, *277*, 3809–3817. [[CrossRef](#)]
32. Rancurel, C.; Khosravi, M.; Dunker, A.K.; Romero, P.R.; Karlin, D. Overlapping Genes Produce Proteins with Unusual Sequence Properties and Offer Insight into De Novo Protein Creation. *J. Virol.* **2009**, *83*, 10719–10736. [[CrossRef](#)]
33. Sabath, N.; Wagner, A.; Karlin, D. Evolution of Viral Proteins Originated De Novo by Overprinting. *Mol. Biol. Evol.* **2012**, *29*, 3767–3780. [[CrossRef](#)]
34. Krakauer, D.C. Evolutionary principles of genomic compression. *Comments Theor. Biol.* **2002**, *7*, 215–236. [[CrossRef](#)]
35. Elena, S.F.; Carrasco, P.; Darós, J.A.; Sanjuán, R. Mechanisms of genetic robustness. *EMBO Rep.* **2006**, *7*, 168–173. [[CrossRef](#)] [[PubMed](#)]
36. Holmes, E.C. Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol.* **2003**, *11*, 543–546. [[CrossRef](#)] [[PubMed](#)]
37. Lillo, F.; Krakauer, D.C. A statistical analysis of the three-fold evolution of genomic compression through frame overlaps in prokaryotes. *Biol. Direct* **2007**, *2*, 22. [[CrossRef](#)] [[PubMed](#)]
38. Krakauer, D.C.; Plotkin, J.B. Redundancy, antiredundancy, and the robustness of genomes. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1405–1409. [[CrossRef](#)] [[PubMed](#)]
39. Peleg, O.; Kirzhner, V.; Trifonov, E.; Bolshoy, A. Overlapping messages and survivability. *J. Mol. Evol.* **2004**, *59*, 520–527. [[CrossRef](#)]
40. Hughes, A.L.; Westover, K.; da Silva, J.; O'Connor, D.H.; Watkins, D.I. Simultaneous positive and purifying selection on overlapping reading frames of the *tat* and *vpr* genes of simian immunodeficiency virus. *J. Virol.* **2001**, *75*, 7966–7972. [[CrossRef](#)]
41. Guyader, S.; Ducray, D.G. Sequence analysis of potato leafroll virus isolates reveals genetic stability, major evolutionary events and differential selection pressure between overlapping reading frame products. *J. Gen. Virol.* **2002**, *83*, 1799–1807. [[CrossRef](#)]
42. Pagán, I.; Holmes, E.C. Long-term evolution of the *Luteoviridae*: Time scale and mode of virus speciation. *J. Virol.* **2010**, *84*, 6177–6187. [[CrossRef](#)]
43. Schneider, W.L.; Roossinck, M.J. Genetic diversity in RNA viral quasispecies is controlled by host-virus interactions. *J. Virol.* **2001**, *75*, 6566–6571. [[CrossRef](#)]
44. Deem, A.; Keszthelyi, A.; Blackgrove, T.; Vayl, A.; Coffey, B.; Mathur, R.; Chabes, A.; Malkova, A. Break-induced replication is highly inaccurate. *PLoS Biol.* **2011**, *9*, e1000594. [[CrossRef](#)]

45. Hicks, W.M.; Kim, M.; Haber, J.E. Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* **2010**, *329*, 82–85. [[CrossRef](#)] [[PubMed](#)]
46. Fondong, V.N. Geminivirus protein structure and function. *Mol. Plant. Pathol.* **2013**, *14*, 635–649. [[CrossRef](#)] [[PubMed](#)]
47. Luna, A.P.; Romero-Rodríguez, B.; Rosas-Díaz, T.; Cerero, L.; Rodríguez-Negrete, E.A.; Castillo, A.G.; Bejarano, E.R. Characterization of Curtovirus V2 protein, a functional homolog of Begomovirus V2. *Front. Plant Sci.* **2020**, *11*, 835. [[CrossRef](#)] [[PubMed](#)]
48. Lacatus, G.; Sunter, G. The Arabidopsis PEAPOD2 transcription factor interacts with geminivirus AL2 protein and the coat protein promoter. *Virology* **2009**, *392*, 196–202. [[CrossRef](#)]
49. Pagán, I.; Betancourt, M.; de Miguel, J.; Piñero, d.; Fraile, A.; García-Arenal, F. Genomic and biological characterization of chiltepín yellow mosaic virus, a new tymovirus infecting *Capsicum annuum* var. *aviculare* in Mexico. *Arch. Virol.* **2010**, *155*, 675–684. [[CrossRef](#)]
50. Deom, C.M.; Brewer, M.T.; Severns, P.M. Positive selection and intrinsic disorder are associated with multifunctional C4(AC4) proteins and geminivirus diversification. *Sci. Rep.* **2021**, *11*, 11150. [[CrossRef](#)]
51. Fondong, V.N. The ever-expanding role of C4/AC4 in Geminivirus infection: Punching above its weight? *Mol. Plant* **2019**, *12*, 145–147. [[CrossRef](#)]
52. Martin, D.P.; Lefeuvre, P.; Varsani, A.; Hoareau, M.; Semegni, J.-Y.; Dijoux, B.; Vincent, C.; Reynaud, B.; Lett, J.-M. Complex recombination patterns arising during geminivirus coinfections preserve and demarcate biologically important intra-genome interaction networks. *PLoS Pathog.* **2011**, *7*, e1002203. [[CrossRef](#)]
53. Simon-Loriere, E.; Holmes, E.C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **2011**, *9*, 617–626. [[CrossRef](#)]