Plant Molecular Biology **26**: 553–556, 1994.
© 1994 *Kluwer Academic Publishers. Printed in Belgium.*

News and views

Experimental and theoretical definition of geminivirus origin of replication

Gerardo Argüello-Astorga, Luis Herrera-Estrella and Rafael Rivera-Bustamante
Centro de Investigación y de Estudios Avanzados del IPN, Unidad Irapuato, Apdo Postal 629, Irapuato, Gto. Mexico

Received and accepted 13 June 1994

Key words: virus evolution, rolling-circle replication, geminivirus, origin of replication

Abstract

Geminiviruses are plant pathogens that replicate by a rolling-circle mechanism, analogous to that used by several prokaryotic ssDNA replicons. Recent reports provide important progress in understanding the structure and functioning of replication origin from these viruses. We have used these data to propose models for the initiation of replication in dicot- and monocot-infecting geminiviruses.

Geminiviruses (GVs) are a large family of plant viruses with genomes of one or two circular single-stranded DNA molecules, which replicate via dsDNA intermediates in the nuclei of plant cells. Due to their simple genomic organization, and to the fact that their replication and transcription greatly rely on host factors, GVs are an excellent model to study such processes in plants.

Several lines of evidence have shown that GVs replicate by a rolling circle (RC) mechanism. Moreover, it has been suggested that the viral protein essential for the replication of GVs [AL1 in dicot-infecting geminiviruses (D-GVs) and L1′/L1″ in monocot-infecting geminiviruses (M-GVs)] is related to the RC replication initiation proteins of some ssDNA bacterial plasmids [1].

Our understanding of the structure and function of geminiviral replication origin (*ori*), located in the intergenic region (IR), is relatively limited. However, recent reports from several laboratories have provided the first pieces of the puzzle. One of the first clues came from Lazarowitz and coworkers, who recently established that the *ori* of two D-GVs, tomato golden mosaic virus (TGMV) and squash leaf curl virus (SqLCV), contains at least two different elements: (1) a 30 nt conserved sequence with the potential to form a stem-loop structure, and (2) sequences upstream of this conserved element, which function as specificity determinants of replication [2]. The first element, the so-called 'structurally conserved element' (SCE), includes in the loop of the potential hairpin the invariant sequence 5′-TAATATTAC-3′. It has been reported that replication associated proteins from M-GVs (L1′/L1″) produce a nick in this sequence to initiate replication [3].

More recently, Hanley-Bowdoin and coworkers reported the experimental identification of the specific binding sites of the AL1 proteins of two D-GVs, TGMV and bean golden mosaic virus (BGMV) and proposed that geminivirus *ori* has a modular organization [4, 5]. In a more theoretical approach involving a phylogenetic and structural analysis of the IR of 30 GVs, a similar con-

clusion was reached [6]. The data from both approaches can be combined to provide important clues for the definition of the geminivirus origin of replication. Here we shall discuss the implications of these reports to propose a model for the initiation of replication in geminiviruses.

Specific binding sites for AL1 protein

The AL1 proteins show high specificity for their cognate *ori* and cannot replicate heterologous genomes (although there are few reported exceptions). Therefore, a great heterogeneity in the DNA sequences recognized by this protein family is expected. The structural analysis of the IR from 22 D-GVs revealed the presence, in all of them, of a series of sequence elements 8–12 nt in length, which are repeated 3–6 times within the region where the *ori*'s of SqLCV and TGMV were mapped (Fig. 1). The nucleotide sequence of the iterated elements (iterons) is generally virus-specific, but their arrangement (relative position, orientation and spacing) is highly conserved between the members of the major lineages of D-GVs (as defined by the amino acid sequence of AL1 proteins). Based on theoretical considerations and an analysis of published data, we proposed that these iterons are the AL1 binding-specific sites [6]. Our model explains experimental results where the ability for replicative complementation (pseudorecombination) between different D-GVs was explored: viruses that complement in replication have identical iterons (e.g. tomato mottle virus and bean dwarf mosaic virus, ToMoV and BDMV, both have the iteron TATTGGAG), whereas viruses that do not complement have iterons with different sequence (e.g. Brazil and Guatemala isolates of bean golden mosaic virus, BGMV-BZ and -GA; AATCGGTG and AATTGGAG, respectively).

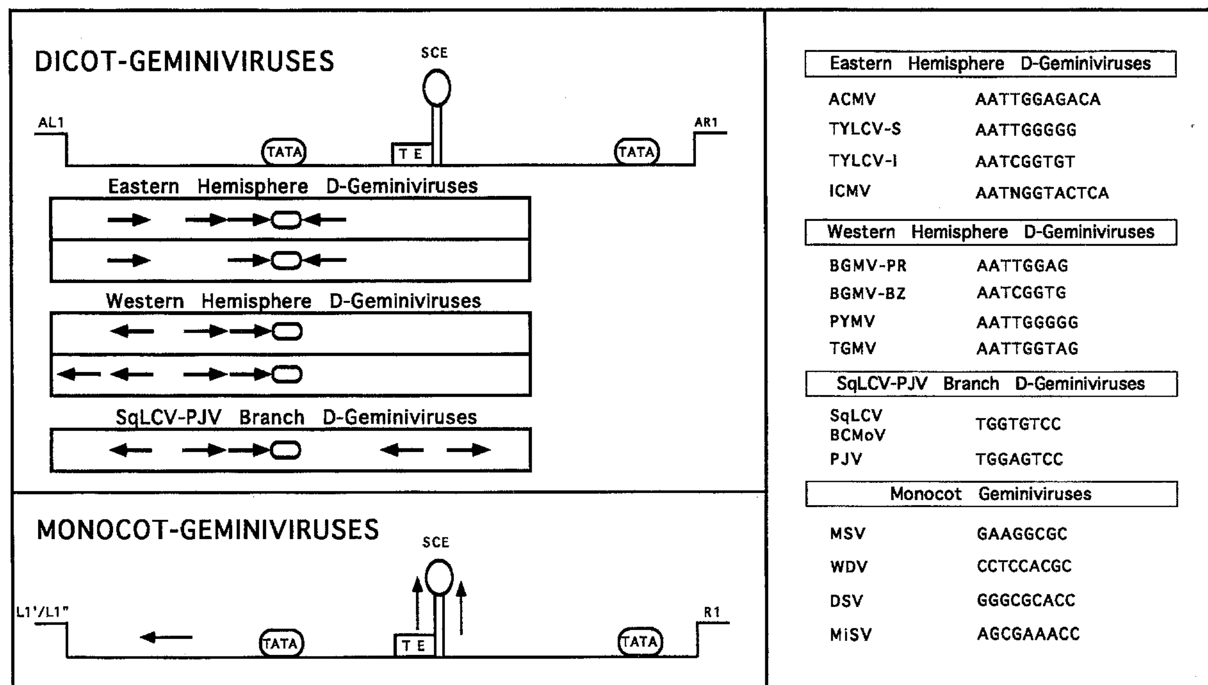

Fig. 1. Arrangement of iterons in different lineage of geminiviruses. Symbols are as follows: arrows, individual iterons; ellipse, putative TATA boxes; TE, transcriptional *cis*-regulatory element; hairpin, the SCE. The consensus sequence of iterons is shown for some geminiviruses of each branch. ACMV, African cassava mosaic virus; BCMoV, bean calico mosaic virus; BGMV-BZ and -PR, Brazil and Puerto Rico isolates of bean golden mosaic virus; DSV, *Digitaria* streak virus; ICMV, Indian cassava mosaic virus; MiSV, *Miscanthus* streak virus; MSV, maize streak virus; PJV, pepper jalapeño virus; PYMV, potato yellow mosaic virus; SqLCV, squash leaf curl virus; TYLCV-I and -S, Israel and Sardinia isolates of tomato yellow leaf curl virus; WDV, wheat dwarf virus.

The reported TGMV and BGMV AL1 protein-binding sites consist of elements that contain two 5 bp direct repeat motifs [4, 5]. When these elements are compared with the iterons that we described for those viruses [6], they correspond to the two conserved iterons found in tandem next to the AL1 TATA box (Fig. 1). Fontes *et al.* also concluded that at least an additional element (besides the SCE and their reported AL1-binding sites) is required for replication, and suggested a modular organization for the replication origin [5]. Whether or not that third element corresponds to one of the extra elements found in our analysis, remains to be tested.

Based on their similarity and the conservation of their arrangement, the iterons found in all D-GVs constitute a series of phylogenetically related elements and, therefore, their function is expected to be equivalent. Thus, experimental data obtained for the functional role of iterons from one virus could be extrapolated to other viruses of the same phylogenetic group. In this way, and supported by the data from Fontes *et al.*, our model establishes simple structural rules for the identification of the AL1-specific binding sites of any D-GV included in the 3 lineages defined until now. This notion has important biological and biotechnological implications in terms of the understanding of geminivirus replication and evolution and for the development of GV-resistant transgenic plants.

Putative binding sites of the M-GV L1′/L1″ protein

Since the nucleotide sequence of the SCE is almost identical among D-GVs, it was initially predicted, and later experimentally confirmed, that this element could not be part of the replication specificity determinants of those viruses [2, 5]. In contrast, the analysis of the IR of M-GVs led to the opposite conclusion for this group of GVs. This conclusion is based in two major considerations: (1) the 'stem' region of the SCE in M-GVs is different for each virus, and (2) a singular structural organization common to all M-GVs examined has been identified ([6]; Fig. 1). The underlying hypothesis is that the 7–9 nt elements present in the stem of the SCE and repeated 25–35 nt downstream of the putative L1′/L1″ gene TATA box, are the specific binding sites of the M-GVs L1′/L1″ protein. Like the D-GVs iterons, the nucleotide sequence of M-GV iterons is virus-specific and has a defined arrangement within the IR. Heyraud *et al.* [7] have recently reported experimental data that supports this hypothesis. These authors mapped the initiation of replication in wheat dwarf virus (WDV) in two distinct sites separated by 170 nt. The first site is located in the SCE and the second one in the 5′ end of a 15 nt sequence highly homologous to the right half of the SCE. Both sites correspond well to the iterons found for WDV in our structural analysis.

Functional organization of geminivirus *ori*

The experimental identification of AL1-specific binding sites in TGMV and BGMV, and the putative equivalent elements predicted by our model for the remaining D-GVs, raises an important question: how is the AL1 protein able to produce a nick in the loop of the SCE, when its specific binding sites lie at a considerable distance away (from 27 to more than 80 nt)? Two alternative models have been proposed in order to answer this question. The first model suggests that the AL1 protein binds to its cognate sequence, melts adjacent dsDNA, translocates along the ssDNA further unwinding the origin until it reaches the invariant sequence in the SCE. There, the AL1 protein produces a nick to initiate plus-strand replication [4]. The second model is suggested by two structural features that we found are common to most D-GVs: (1) the association of iterons with the AL1 gene TATA box, and (2) the intimate association of lineage-specific conserved sequences with the 5′ end of SCE. These sequences are very similar to well-known plant *cis*-regulatory elements (a G-box-like element in American D-GVs, and an AT-rich element in Eurasian-African D-GVs). In this model, a host transcription factor bound near the SCE interacts with a TATA-binding protein (TBP) or other

TBP-associated factor. This interaction forces the looping out of the intervening DNA, thus bringing together the SCE and the AL1 proteins bound to the TATA-proximal iterons ([6]; Fig. 2).

Our structural analysis suggests a different organization of M-GVs *ori*. In this case, considering the close association of the putative L1′/L1″ binding and cleavage sites (the invariant sequence), it can be expected that the L1′/L1″ molecule bound to its cognate site, is also in an adequate position to perform the nick on the viral DNA. Conserved sequences lying next to the SCE borders, as the well-characterized *rpe1* transcriptional activator element of maize streak virus (MSV) [8], could play an auxiliary (essential?) role in the viral replicative process, by binding host transcription factors whose main function could be the creation of a nucleosome-free region in the vicinity of the SCE to allow the access of the L1′/L1″ protein to their cognate sequences within this element.

Although our model is based mostly on theoretical considerations and is highly speculative, it is consistent with the experimental data published to date and can be experimentally tested.

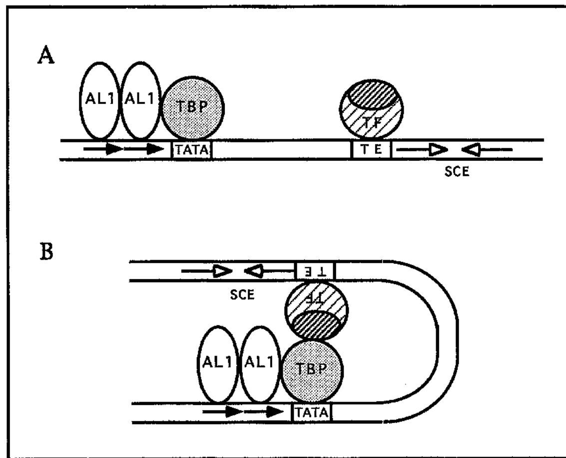

Fig. 2. Hypothetical model for the functional organization of the replication origin in dicot geminiviruses (see text). TBP, TATA-binding protein; TE, transcriptional *cis*-regulatory element; TF, host transcription factor; SCE, putative stem-loop structure; filled arrows: iterons.

References

1. Koonin EV, Ilyina TV: J Gen Virol 73: 2763–2766 (1992).
2. Lazarowitz SG, Wu LC, Rogers SG, Elmer JS: Plant Cell 4: 799–809 (1992).
3. Heyraud F, Matzeit V, Schaefer S, Schell J, Gronenborn B: Biochimie 75: 605–615 (1993).
4. Fontes EPB, Eagle PA, Sipe PS, Luckow VA, Hanley-Bowdoin L: J Biol Chem 269: 8459–8465 (1994).
5. Fontes EPB, Gladfelter HJ, Shaffer RL, Petty ITD, Hanley-Bowdoin L: Plant Cell 6: 405–416 (1994).
6. Argüello-Astorga GR, Guevara-Gonzalez RG, Herrera-Estrella LR, Rivera-Bustamante RF: Virology 203: 90–100 (1994).
7. Heyraud F, Matzeit V, Kammann M, Schaefer S, Schell J, Gronenborn B: EMBO J 12: 4445–4452 (1993).
8. Fenoll C, Schwarz JJ, Black DM, Schneider M, Howell SH: Plant Mol Biol 15: 865–877 (1990).