Volume 16 Number 14 1988 Nucleic Acids Research

Genetic analysis of the tomato golden mosaic virus II. The product of the AL1 coding sequence is required for replication

J.Scott Elmer, Leslie Brand, Garry Sunter[1], William E.Gardiner[1], David M.Bisaro[1] and Stephen G.Rogers*

Plant Molecular Biology, Biological Sciences, Monsanto Company, St Louis, MO 63198 and [1]Department of Molecular Genetics and Ohio State Biotechnology Center, Rightmire Hall, Ohio State University, Columbus, OH 43210, USA

Received December 31, 1987; Revised and Accepted May 17, 1988

ABSTRACT

Tomato golden mosaic virus (TGMV) belongs to the geminivirus subgroup that is characterized by a split genome consisting of two single-stranded circular DNAs. The TGMV A genome component encodes the virus coat protein as well as all of the functions necessary for viral DNA replication. Analysis of the nucleotide sequence indicates that the TGMV A component has, in addition to the coat protein encoding ORF, four overlapping open reading frames (ORFs) with the potential to encode proteins of greater than 10 kD. We have investigated the functions of these putative proteins in both symptom formation and DNA replication by creating mutations in each of the ORFs. Our results show that the AL4 ORF, which is encoded within the N-terminal region of ORF AL1, is not essential for normal virus infection. In contrast, we find that disruption of the AL3 ORF results in delay and attenuation of symptom formation. We also report that the products of the AL1 and AL2 ORFs are absolutely required for symptom formation. Studies of DNA replication show that only the AL1 open reading frame is essential for viral DNA synthesis. The significance of these results for the development of vectors from the geminiviruses is discussed.

INTRODUCTION

The geminiviruses are unique among viruses that infect eucaryotes in possessing small, twinned isometric particles containing single-stranded circular DNA genomes. Two subgroups of geminiviruses can be distinguished by both their insect vector and their genome content. Several recent reviews[1,2,3] of the geminiviruses are cited in the references. One subgroup is typified by maize streak virus (MSV)[4,5], wheat dwarf virus (WDV)[6] and beet curly top virus (BCTV)[7] which are leafhopper-transmitted and contain a unique single-stranded circular DNA genome. Viruses of the second subgroup, typified by cassava latent virus (CLV)[8] and tomato golden mosaic virus (TGMV)[9], are whitefly-transmitted and contain a genome split between two single-stranded circular DNAs. Viruses of the second subgroup and their DNAs are mechanically transmissible and their cloned DNAs are infectious in combination. The nucleotide sequences of CLV[10] and TGMV[11] have been determined and show that the two components of each of these viruses share nearly identical sequence over a region of 230 to 250 base pairs (Figure 1). It has been proposed that this region of

major homology, referred to as the common region, may function as the origin of replication although experimental demonstration of this hypothesis remains to be provided.

Comparison of the computer-identified open reading frames (ORFs) of the geminiviruses reveals a high degree of homology and similar organization among the various two component geminiviruses[5,6]. One of the ORFs encodes the viral coat protein[12], which has recently been shown not to be essential for CLV infectivity[13]. Our own studies showed that deletion of the TGMV coat protein coding sequence results in attenuated symptom development and delayed systemic spread of the virus but, likewise, has no effect on replication[14].

In addition to the coat protein, the TGMV A component encodes four other overlapping open reading frames (ORFs) with the potential to encode proteins of greater than 10 kD. One of these ORFs, AL1 (Figure 1), is remarkably similar to two ORFs of the single component geminiviruses, which suggests that these two ORFs have either become fused or separated during evolution. The sequence conservation has also led to speculation that the product of AL1 is necessary for replication. Indirect support for this hypothesis comes from the recent work of Townsend et al.[15] who demonstrated by transfection of tobacco protoplasts with cloned DNAs that the CLV 1

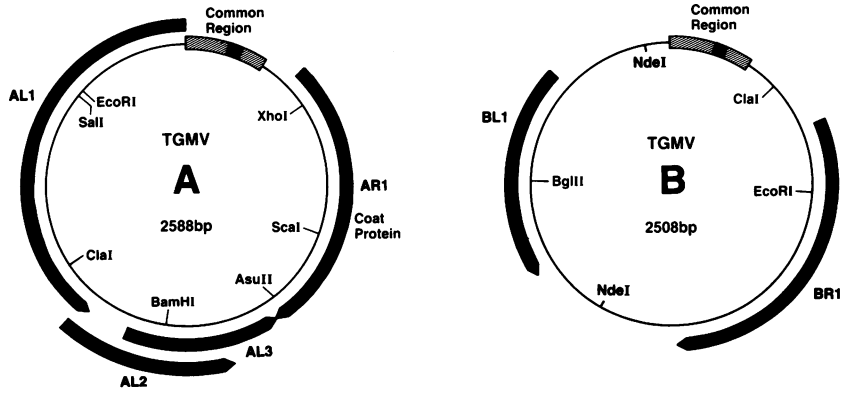

Figure 1. Maps of the TGMV double-stranded DNAs. The positions and extents of the open reading frames (ORFs) encoded by the TGMV A and B genome components are shown as solid arrows. The Common Region designates 230 bp of nearly identical sequence shared by the two DNAs; the solid box in the common region represents the stem and loop structure found in the sequences of all geminiviruses[3]. The left sides of the common region are the origins of the A and B sequences[11] which then proceed clockwise around the maps. The standard three letter designations for restriction endonuclease cleavage sites have been used[32]. The ClaI sites are unique in DNA prepared from dam+ DNA adenine methylase-proficient E. coli cells.

component (analogous to the TGMV A component) contains all the information necessary for its own DNA replication. Similar findings have been obtained by Rogers et al.[16] who showed that transgenic petunia plants containing integrated dimers of the TGMV A component produce free, autonomously replicating TGMV A DNAs. Both of these studies clearly show that component A, which encodes the highly conserved AL1 ORF, produces all the virally-encoded factors necessary for its own replication. It is also clear that these factors are likely to be proteins since they act in trans to promote replication of TGMV B.

We have used mutations created in vitro to investigate the roles of the products encoded by the four overlapping TGMV A ORFs in symptom formation and replication. We present evidence that the AL4 ORF is not essential for normal virus infection. In contrast, we find that disruption of the AL3 ORF results in delay and attenuation of symptom formation and that the products of the AL1 and AL2 ORFs are absolutely required for symptom formation. Most importantly, analysis of replication by these mutants showed that only the product of the AL1 open reading frame is essential for viral DNA replication.

MATERIALS AND METHODS

Geminivirus nomenclature

To standardize nomenclature for the geminivirus ORFs, we have followed the convention of designating the ORFs using their positions on the standard map[11] (Figure 1) rather than by the less descriptive method of referring to the ORFs by the size of the putative protein produced[8]. All of the leftward ORFs are translated from the strand opposite to the standard sequence and are delineated using higher to lower numbers. The AL1 ORF is an exception to this rule since AL1 crosses the sequence origin (defined as the left side of the common region sequences[11]) of the 2588 nucleotide TGMV A component. ORF AL1 is encoded by nucleotides 13 to nucleotide 0 and continues through nucleotides 2588 to 1545, AL2 from nucleotides 1601 and 1214, ORF AL3 from nucleotides 1465 and 1069 and ORF AL4 from nucleotides 2447 and 2186.

Construction and molecular description of the mutations created near or in the TGMV A leftward ORFs

Restriction endonucleases, DNA modifying enzymes and synthetic linkers were obtained from New England Biolabs (Beverly, MA) or Bethesda Research Laboratories (Gaithersburg, MD) and were used according to the manufacturer's directions. DNA fragments were purified and reassembled as described previously[16]. All sequence alterations in the mutants were indirectly confirmed by restriction endonuclease digestion if new cleavage sites were created, or directly demonstrated by di-deoxy

DNA sequencing of single-stranded templates[17] or double-stranded plasmids[18,19].

The construction of pMON336 and pMON351, the binary Agrobacterium tumefaciens Ti plasmid based vectors for introduction of TGMV A DNA into permissive host plants has been described previously[20]. Briefly, pMON336 is a derivative of pMON505[21] that carries the 1.5 kb EcoRI to ScaI fragment of TGMV A derived from pMON344[20]. The pMON351 plasmid is a derivative of pMON505 that carries the 1.1 kb EcoRI to ScaI fragment of TGMV A derived from pMON344. Both plasmids retain a unique EcoRI site for insertion of a full length TGMV A DNA to create greater than unit length copies of TGMV for Agrobacterium tumefaciens delivery to plants. The full length TGMV A insert in pMON336 gives pMON341 and the full length insert in pMON351 gives pMON337. Agrobacterium tumefaciens carrying these plasmids were the wild-type controls for the mutants described below.

In the descriptions of the TGMV A leftward ORFs that follow, the numbers refer to the nucleotide coordinates of the TGMV A sequence determined by Hamilton et al.[11]. All of the TGMV A leftward ORF mutations are diagrammed in Figure 2. The proteins described below are predicted from the sequence and have not been physically identified.

Mutation AL1-1 was created by site directed mutagenesis of TGMV A DNA inserted as an EcoRI fragment in pUC119 (provided by J. Messing, Waksman Institute) using the procedure of Kunkel[22]. Plasmid pUC119 is a derivative of pUC19[23] that may be isolated as single-stranded DNA after superinfection with M13K07, a variant of M13 impaired in its ability to replicate. The mutagenesis resulted in the following sequence (inserted bases are in brackets):

(n. 30) 5' -GCC[AGA]TCTTAATTACAAAA[CAT]ATG (formerly n. 11)

This mutagenesis resulted in the insertion of AGA to create a BglII site and CAT to create an NdeI cleavage site 5' to the AL1 ORF. The initiator ATG of the AL1 ORF is part of the new NdeI (5'-CATATG) site. There is no alteration of the presumed AL1 ORF coding sequence. The mutated pUC119 clone is pMON434. The EcoRI fragment was transferred into the EcoRI cleaved binary Ti vector, pMON336, to give one and one half copies of the TGMV A DNA but only a single copy of the altered DNA sequence. The resulting binary vector is pMON481.

Mutation AL1-2 was created by digestion of pMON337 DNA with ClaI which cleaves at nucleotide 1814 within the AL1 ORF coding sequence, followed by treatment with the large fragment of DNA polymerase I and religation. A plasmid, pMON421 was isolated which lacks the ClaI site but does not contain the predicted new NruI site. Direct sequencing of the pMON421 DNA using a synthetic primer revealed that a 12 nucleotide deletion had occurred as shown in brackets below.

(n. 1801) 5'-CGGTGTGAC[ATCATCGATGAC]GTTATA (former n. 1827)

This deletion resulted in the in-frame removal of the codons for four amino acids located 88 amino acids from the C terminus of the AL1 protein. The derivative of pMON351 carrying this deletion is pMON421.

Mutation AL1-3 was created by site directed mutagenesis to delete a single A at nucleotide 1602 to create an NdeI site at the ATG of the AL2 coding sequence in the sequence shown below. The deleted A appears in brackets.

(n. 1610) CTTTCCAT[A]ATG

Since the AL2 and AL1 ORFs overlap, this deletion resulted in a single base frameshift mutation in the the AL1 coding sequence. The sequence of the mutation predicts replacement of the 19 C-terminal amino acids of the AL1 protein with the entire AL2 protein sequence. The pUC119 TGMV A EcoRI insert derivative carrying the mutation is called pMON432. The pMON351 binary vector carrying this mutated EcoRI fragment is pMON446.

Mutation AL1,4-1 was created by inserting the linear EcoRI TGMV A fragment into a derivative of pAT153 from which the SalI site had been removed. This plasmid was then digested with SalI which cleaves the TGMV A insert at a single site at nucleotide 2242. The SalI site is located within the coding sequences for both the AL1 and AL4 ORFs. After treatment with Klenow polymerase and ligation, a plasmid lacking the SalI site but carrying a new PvuI site was identified and called pTGA8. The plasmid contains a 4 nucleotide insertion into both the AL1 and AL4 ORFs creating frameshift mutations in both. The AL1 protein is predicted to terminate after the addition of five amino acids resulting in a truncated protein consisting of the N terminal 33 percent of the AL1 protein. The frameshift is also predicted to fuse the N terminal 82 percent of the AL4 protein to a 67 percent C terminal fragment of the AL1 protein. The mutated EcoRI fragment was transferred to pMON351 to give plasmid pMON441.

Mutation AL4-1 was created by site directed mutagenesis to introduce a termination codon into the AL4 coding sequence and a silent mutation into the AL1 coding sequence. The resulting sequence is shown below where the A in brackets is a T in the wild type sequence.

(n. 2420) 5'-TGCT[A]ATTCAG (n. 2410)

This nucleotide change results in the production of the wild type AL1 protein and would produce a truncated AL4 protein consisting of 10 percent of its N-terminus. The pUC119 derivative plasmid carrying the mutation is pMON442. The pMON336 derivative carrying the mutated EcoRI fragment is pMON452.

Mutation AL2-1 was created by insertion of a C residue to introduce an XbaI site into the AL2 coding sequence just beyond the end of the AL1 coding sequence. In the sequence shown below, the inserted C is in square brackets. The XbaI sequence contains a translational termination signal TAG which is in frame with the AL2 coding sequence.

(former n. 1540) 5'-GAAGAGCTAT[C]TAGAAGGCGAC (n. 1520)

This mutation is predicted to produce a truncated AL2 protein consisting of the N terminal 23 amino acids (18 percent). The pUC119 TGMV A EcoRI insert derivative carrying the mutation is called pMON427. A derivative of pMON351 carrying the mutated EcoRI fragment is called pMON422.

Mutation AL3-1 was created using site directed mutagenesis to delete 88 nucleotides encoding the C terminal region of the AL3 coding sequence and to insert an XbaI site into the AL3 coding sequence immediately adjacent to the end of the AL2 coding sequence. The putative polyadenylation signal sequence for the leftward transcripts is not affected by this deletion. In the resulting sequence, shown below, the nucleotides adjacent to the deletion are in square brackets with three newly inserted nucleotides between them. The XbaI sequence contains a translational termination signal TAG that is in frame with the AL3 coding sequence.

(n. 1111) 5'-TTATATATGATC[T]AGA[C]TATTTAAA (former n. 1220)

This mutation is predicted to produce an N terminal fragment of 83 amino acids or 63 percent of the AL3 protein. The XbaI site was created adjacent to a 5'-GATC dam methylation site and can only be cleaved in DNA prepared from a dam-deficient cell such as GM48[24]. The pUC119 TGMV A EcoRI insert derivative carrying the mutation is called pMON428. Plasmid pMON438 is the derivative of pMON351 carrying the mutated EcoRI fragment.

Mutation AL2,3-1 was created after transfer of the TGMV A EcoRI fragment into a derivative of pAT153 with no BamHI site. This plasmid was then digested with BamHI which cleaves the TGMV A DNA at a single site at nucleotide 1356, the ends repaired by treatment with the Klenow fragment of DNA polymerase and religated. A plasmid lacking the BamHI site and containing a new ClaI site, as determined by digestion of DNA prepared from GM48 cells, was identified and called pTGA3. This four base insertion results in frameshift mutations in both the AL2 and AL3 coding sequences. The sequence predicts that the AL2 ORF will end after the addition of 21 amino acids from the third reading frame in this sequence to the N terminal 79 amino acids of the AL2 protein. The altered sequence also predicts a hybrid AL3-AL2 protein with the N-terminal 36 amino acids of the AL3 protein fused to the C terminal 48 amino acids of the AL2 protein. The EcoRI fragment carrying the mutation was transferred into pMON351 to give plasmid pTGA5.

Mutation AL2,3-2 was created by site directed mutagenesis of pMON420, which carries the XbaI site located in AL2, to introduce the deletion and additional XbaI site as described above for mutation AL3-1. DNA from the mutated plasmid was prepared from dam-deficient cells, cleaved with XbaI and rejoined. In the final pUC119 construction, 400 bp located between the XbaI sites was deleted. This deletion is

predicted to produce a truncated AL2 protein consisting of the N terminal 23 amino acids (18 percent) of the protein and no AL3 protein. The pUC119 TGMV A EcoRI insert derivative carrying the deletion is called pMON436. The EcoRI fragment was transferred into pMON351 to produce pMON439.

Assays for symptom production

Infectivity assays were performed as described by Elmer et al.[20] using transgenic Nicotiana benthamiana plants carrying chromosomally integrated tandem copies of the TGMV B DNA as hosts. The plants were either mechanically inoculated with 20 μg per plant of the linear, mutant TGMV A DNA released by EcoRI digestion of plasmid DNA or inoculated using a more efficient agroinoculation procedure, in which the transgenic plants were inoculated with A. tumefaciens carrying a greater than unit length copy of the mutant TGMV A DNA in a binary transformation vector.

Analysis of replication in infected plants

Leaf tissue was taken from plants for DNA analysis when the newly emerged leaves showing chlorotic lesions became fully expanded, which generally occurred between 3 to 4 weeks after inoculation depending on the mutant and delay of symptom formation. Leaves were taken from at comparable times and positions from plants inoculated with mutants that did not produce symptoms. Total leaf DNA isolation, nick translation of DNA probes and Southern blot hybridization analysis were carried out using the component-specific probes, either the 1460 bp ScaI to EcoRI fragment of component A in pMON349 or the 1570 bp ClaI to BglII fragment of component B in pMON350 as described previously[16].

Analysis of replication in leaf discs

Leaf discs, 6 mm (1/4 inch) in diameter, were prepared from surface sterilized wild type Mitchell diploid petunia leaves. Following culture on MS104 agar medium for 2 days to promote partial cell wall formation at the wound surfaces, the discs were submerged and shaken gently for three minutes in a saturated culture of A. tumefaciens cells containing the disarmed pTiB6S3-SE plasmid and binary vectors carrying one and one-half copies of the mutated TGMV A DNA. The discs were removed from the bacterial suspension, blotted dry, and incubated upside down on MS104 agar medium. After 2 days, the discs were transferred to selection media containing MS media with 500 μg/ml carbenicillin and 300 μg/ml kanamycin. Total DNA was isolated from six leaf discs (approximately 0.5 g) using the procedure of Dellaporta et al.[25], 4 to 8 days after transfer to selection media. Uncut DNA was separated by electrophoresis in 0.8% agarose gel, blotted onto nitrocellulose, and hybridized to a nick translated probe as described above.

Biosafety Considerations

All of the experiments described were conducted according to the NIH

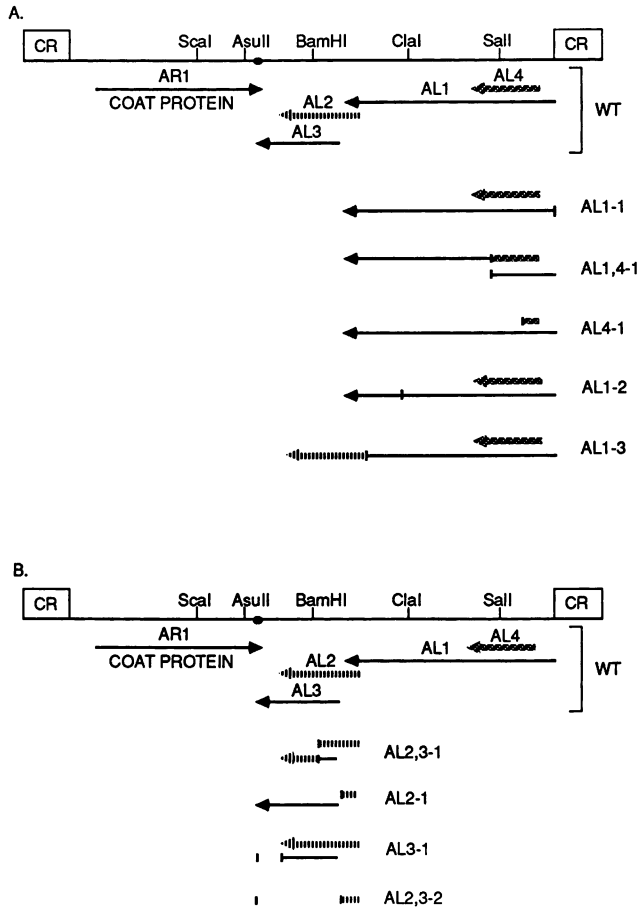

Figure 2. Diagrams of the ORFs in wildtype and mutants of the TGMV A component. The predicted proteins produced by each of the ORFs in the wildtype virus (WT) and mutants are shown as arrows for wild-type or fusion proteins or as lines when the ORF is truncated. Vertical bars mark the positions of the mutations and endpoints of truncated ORFs or endpoints of deletion mutations. A. shows the results of mutations in the AL1 and AL4 ORFs; B. shows the results of mutations in the AL2 and AL3 ORFs. Detailed descriptions of these mutations appear in the text. CR is the common region which is repeated in this linear representation of the circular TGMV A DNA. The restriction endonuclease cleavage sites shown are unique except for the ClaI site.

Guidelines for Recombinant DNA Research. Transgenic plants with tandem repeats of the geminivirus components in their genomes were propagated in limited access growth chambers. All plants and soil were sealed in plastic bags and autoclaved before disposal. For experiments involving Agrobacterium inoculation, the runoff from

Table 1. Time of Appearance of Virus Symptoms After Agrobacterium Stem Inoculation with Various Mutants

Mutation	Inoculum[a]	No. expts.	No. inf./ No. inoc.	Days until Symptoms Range	Avg.	Std.dev.
wild type	337-SE	13	140/143	7-19	12.3	2.5
AL1-2	421-SE	1	0/18	-	-	-
AL1-3	446-SE	2	0/15	-	-	-
AL2-1	422-SE	2	0/29	-	-	-
AL3-1	438-SE	3	36/36[b]	19-28	23.3	1.9
AL2,3-1	pTGA5-SE	2	1/30[c]	-	39[c]	-
AL2,3-2	439-SE	2	0/29	-	-	-
wild type	341-SE	5	60/62	10-22	15.3	3.2
AL1-1	481-SE	2	24/26	11-18	14.3	2.0
AL1,4-1	441-SE	2	0/24	-	-	-
AL4-1	452-SE	2	38/40	12-23	16.4	2.6

[a] The data was obtained from two or more separate experiments. The appropriate control, either pMON337 or pMON341, was tested in parallel with each of the mutants.

[b] The symptoms produced by the AL3-1 mutation were attenuated and consisted of vein clearing without leaf curling.

[c] This single infected plant displayed attenuated symptoms characteristic of the AL3-1 mutation and is presumably a revertant due to a second frameshift mutation that restored AL2 protein function.

watering was collected and treated with hypochlorite (Chlorox, 1 percent) before disposal to kill any bacteria that might wash off the inoculated plants. No Agrobacterium strains were created that carried more than one component (approximately 50 percent of the TGMV genome) in accordance with the NIH Guidelines for cloning of virus genomes.

RESULTS

Construction and assay of TGMV A leftward ORF mutations

To clearly define the role of the putative proteins produced from the leftward ORFs in the infection process, we constructed mutants carrying lesions in the coding sequences of these ORFs. The details of their construction can be found in Material and Methods. A diagrammatic representation of these mutations appears in Figure 2. The proteins described below are predicted from the sequence and have not been physically identified.

The modified TGMV A components were inserted into an appropriate A.

tumefaciens binary T-DNA vector and introduced into complementing Nicotiana benthamiana plants with tandem copies of the TGMV B component integrated into their genomes as previously described[20]. Symptom appearance was recorded visually each day beginning 5 to 6 days after inoculation. The results of these assays are summarized in Table 1. Note that the two controls (337-SE and 341-SE) display different times of symptom appearance. We have proposed as a possible explanation for this difference that the presence of two common regions in pMON337 permits the generation of circular forms by replication while the single common region in pMON341 requires recombination for release[20]. Since the various mutations are located in different places in the TGMV A sequence, the 337-SE strain is the appropriate control for some mutants while the 341-SE strain is appropriate for others as shown in Table 1. All mutations were tested in the plasmid configuration with a single copy of the altered sequence.

Effects of the leftward ORF mutations on symptom formation

The products of ORFs AL1 and AL2 are required for symptom formation Four separate mutations were created in or near the AL1 ORF and three in the AL2 ORF. Mutation AL1-1 creates a BglII and an NdeI cleavage site 5' to AL1 ORF at the ATG translational initiator but does not alter the presumed coding sequence. This mutation had no affect on symptom formation. In contrast, all three mutations within the AL1 coding sequence abolished symptom formation. These mutations were created in both the N-terminal and C-terminal regions of the presumptive AL1 protein. AL1,4-1 is a frameshift mutation positioned one-third of the way into the AL1 coding sequence. Mutations AL1-2 and AL1-3 are frameshift mutations located 88 and 19 amino acids, respectively, from the C-terminus. Since all three of these AL1 mutants prevented symptom formation, we conclude that the AL1 protein is essential for normal infection.

All three mutations in the AL2 ORF also abolished symptom formation. Mutation AL2-1 creates an amber mutation in the AL2 ORF near the N-terminus. Mutation AL2,3-1 creates frameshift mutations in both the AL2 and AL3 ORFs while mutation AL2,3-2 results in the deletion of most of the AL2 and AL3 ORFs. Plants inoculated with these mutants displayed no symptoms. Symptoms that appeared on a single plant inoculated with the AL2,3-1 mutant were greatly delayed and attenuated compared to wild type controls. These symptoms were identical to those produced by the AL3-1 mutant described below and, consequently, this plant most likely carried a revertant that restored AL2 function. Unfortunately, attempts to passage and recover this presumed revertant were unsuccessful. From these results and those described below for the AL3 mutant, we conclude that the AL2 protein is essential for symptom formation.

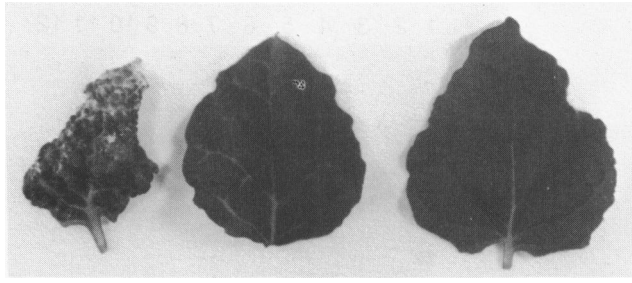

Figure 3. Symptomatic leaves from wild type and mutant infected complementing tobacco plants. The smaller leaf on the left was produced after wild type TGMV A infection. Note the extreme curling and chlorosis. The leaf in the middle was produced by infection with the AL3-1 mutant and displays little curling and only discontinuous vein clearing. The leaf on the right is from an uninoculated control plant.

Deletions in the AL3 ORF result in delayed appearance and attenuation of symptoms

The requirement for the AL3 ORF protein was determined by deleting sequences located near its C-terminus. Plants inoculated with mutant AL3-1 showed symptoms after 24 days compared to an average of 13 days in plants inoculated with the wild type TGMV A 337-SE control. The symptoms induced by the mutant were greatly attenuated compared to those produced by wild type TGMV A infection (Figure 3). The AL3-1 infected leaves did not display the rugosity and extreme chlorosis characteristic of TGMV A infection but, instead, were characterized by discontinuous vein clearing with chlorosis confined to small regions over the veins. From this result we conclude that the protein produced by the AL3 ORF is not essential for symptom appearance but modulates the timing and severity of these symptoms.

The function of the AL4 ORF is dispensable for virus replication and spread

The AL4 ORF is contained entirely within the AL1 ORF but is encoded by another reading frame. To separately determine the importance of the AL4 ORF in symptom formation, mutation AL4-1 was created by site directed mutagenesis which introduced an amber mutation near the N-terminus of this protein while creating a silent, third position change in the AL1 ORF. Complementing TGMV B plants inoculated with this mutant showed normal time of appearance and severity of symptom development. From this result we conclude that the putative AL4 protein is not required for normal symptom formation.

Analysis of DNA forms produced by the mutants in complementing plants

DNA prepared from complementing plants inoculated with A. tumefaciens carrying the various mutants was analyzed by southern blot hybridization using a TGMV A specific

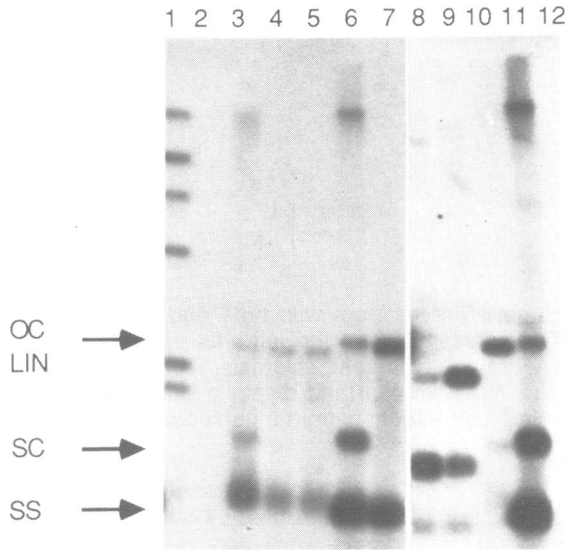

Figure 4. Southern blot analysis of TGMV A DNAs produced in complementing tobacco plants and petunia leaf discs. Hybridizations were carried out as described under Materials and Methods using a nick-translated component A specific probe. Total DNA from leaves of inoculated plants (Lanes 2-7 and 11) or from leaf discs (Lanes 8-10 and 12) was prepared as described under Materials and Methods. Each lane contains 2-5 μg of DNA except Lane 1 which contains ^{32}P-labeled HindIII digested Λ DNA marker fragments. Lane 2, total leaf DNA from mutant AL1-2 agroinoculated complementing plants; Lane 3, total leaf DNA from mutant AL3-1 inoculated complementing plants; Lane 4, total leaf DNA from mutant AL3-1 agroinoculated complementing plants digested with EcoRI; Lane 5, total leaf DNA from mutant AL3-1 agroinoculated complementing plants digested with XbaI; Lane 6, total leaf DNA from mutant AL4-1 agroinoculated complementing plants; Lane 7, total leaf DNA from mutant AL4-1 agroinoculated complementing plants digested with EcoRI; Lane 8, total DNA prepared from mutant AL2,3-2 agroinoculated leaf discs; Lane 9, total DNA prepared from mutant AL2,3-2 agroinoculated leaf discs digested with XbaI; Lane 10, total DNA prepared from wild type TGMV A agroinoculated leaf discs digested with EcoRI; Lane 11, total leaf DNA from wild type TGMV A agroinoculated complementing plants; Lane 12, total DNA prepared from mutant AL1-2 agroinoculated leaf discs. Lanes 1-7 were from a one day exposure at room temperature; lanes 8-12 were exposed for four days at room temperature.

probe as described in Materials and Methods. The results shown in Figure 4 demonstrate that the AL4-1 (Lane 6) and AL3-1 (Lane 3) mutants produce all of the DNA forms characteristic of wild type TGMV A infection (Lane 11). The quantities and proportions of the various forms are similar to those seen in DNAs prepared from wild-type TGMV A infected plants. From this result, we conclude that the AL3 and AL4

mutants are unaffected in replication and systemic spread during infection of whole complementing plants. In contrast, the defects in the AL1 and AL2 mutants prevent replication and/or systemic spread of the virus, since total leaf DNA prepared from AL1 mutant (Lane 2) and AL2 mutant (data not shown) inoculated plants contains none of the forms even when prepared in parallel with the AL3 mutant DNAs (30 days post inoculation).

The DNA forms detected in the symptomatic inoculated plants are produced by the mutants and are not revertants. This is especially clear for the AL3-1 mutant which contains a new XbaI site and a deletion of approximately 80 nucleotides. Undigested (Lane 3), EcoRI-digested (Lane 4) and XbaI digested (Lane 5) DNA from AL3-1 infected leaves have open circular (OC) and linear (LIN) forms that are smaller than the open circular (Lane 6) and the EcoRI-digested linear forms (Lane 7) of full length AL4-1 TGMV A DNA. From this result, we conclude that the AL3-1 mutant retains the deletion and the new XbaI site. Although there is no restriction endonuclease site with which to verify the presence of the AL4-1 mutation, the lack of delay in wild-type symptoms produced by this mutant allowed us to conclude that the AL4-1 DNA forms have been produced by the mutant and not by a revertant, since revertants would be expected to show some delay and/or attenuation.

Analysis of DNA forms produced by the mutants in leaf discs

We have developed a convenient, short term assay for TGMV A replication in normal petunia leaf discs inoculated with A. tumefaciens carrying greater than unit length copies of the TGMV A DNAs. We used this assay to analyze replication, independent of symptom formation and systemic spread, in leaf discs agroinoculated with the mutations in the leftward ORFs. Total DNA was prepared from leaf discs 6-10 days after cocultivation with the Agrobacterium strain, separated by electrophoresis, transferred to nitrocellulose and hybridized with a TGMV A specific probe. The DNA forms produced by mutant AL2,3-2, which has the 400 nucleotide deletion of the AL2 and AL3 ORFs, are shown in Figure 4, Lanes 8 (undigested) and 9 (partial XbaI digestion). All three forms characteristic of TGMV A infection are present, i. e. the open circular, the supercoiled and single-stranded forms. Although the production of single-stranded DNAs in leaf discs is variable and low, this form is clearly present and sensitive to digestion with mungbean nuclease (data not shown). The 400 nucleotide deletion is evident from the sizes of all three forms as compared to EcoRI-linearized wild type TGMV A prepared from leaf discs (Lane 10) and untreated wild type TGMV A DNA prepared from infected complementing plants (Lane 11). From this result and similar results obtained with mutant AL 2,3-1 (not shown), we conclude that the products of the AL2 and the AL3 ORFs are not required for production of the single- and double-stranded viral DNA forms. In contrast, none of the TGMV A DNA forms

were detected following inoculation with AL1-2 mutants (Lane 12). Analysis of DNA from discs inoculated with the AL1,4-1 and the AL1-3 mutants also showed that these DNAs were absent. We conclude from these results that the protein produced by the AL1 ORF is essential for the formation of the single- and double-stranded DNA forms characteristic of TGMV A infection.

DISCUSSION

The experiments described in this report provide direct evidence that the AL1, AL2 and AL3 ORFs of TGMV A DNA are genes whose products are required for normal infection and symptom production. The phenotype of a TGMV A mutant carrying an AL3 C-terminal proximal deletion is one of delayed and attenuated symptom formation although the viral DNA forms found in normal infection are produced. Similar free DNA forms are not present in the symptomless, complementing TGMV B plants inoculated with AL1 or AL2 mutants. We conclude that mutations in either the AL1 or AL2 ORFs totally abolish systemic spread and symptom production. These findings for the AL1 and AL2 mutations are in agreement with the results of Brough et al. [33] who showed that mutations affecting either the AL1 or both the AL2 and the AL3 ORFs abolish infectivity. Since these authors did not create individual mutations in the AL2 and AL3 ORFs, they were unable to separate the functions of the AL2 and AL3 ORFs as our experiments have done.

In contrast to the other computer identified leftward ORFs, the product of the AL4 ORF is not required for normal symptom production and systemic spread of the virus. An amber mutation in the putative AL4 ORF, which is contained within the N-terminal portion of the highly conserved AL1 protein, has no affect on time of symptom appearance or symptom severity. This result suggests that the AL4 ORF and the ORFs found at similar positions in CLV 1[11], BGMV 1[26] and BCTV[7] do not encode essential proteins. However, the possibility that the AL4 may be involved in other aspects of the TGMV life-cycle such as whitefly transmission cannot be excluded by the experiments reported here.

To determine if these mutations had a direct affect on viral DNA replication, we examined DNA found in normal petunia leaf discs after the introduction of the TGMV A mutants using an Agrobacterium Ti plasmid vector. The results show that deletion of most of the AL2 and AL3 ORFs has no effect on formation of single- and double-stranded circular DNAs. Therefore, the products of the AL2 and AL3 ORFs are not involved in replication.

Mutations in the AL1 gene totally abolish viral DNA replication. No single- or double-stranded DNA forms can be detected in DNA from leaf discs infected with these mutants. The AL1 ORF is highly conserved in both the single and two

component genome geminiviruses [6,27]. In the single genome geminiviruses, this ORF is divided into two separate ORFs. Although our AL1 mutations cover the entire length of the coding sequence, we have not identified the proteins the mutant viruses produce and they could still produce an N-terminal fragment. Therefore, we can conclude that the function of, at least, the C-terminal ORF in the single component geminiviruses is likely required for replication of the single component genomes.

The geminiviruses are unique eucaryotic viruses because of their single-stranded circular DNA genomes and this genome structure most resembles that of the coliphages ϕX174 and fd (and closely related phage M13). In this regard, the AL1 gene is analogous to the ϕX174 gene A or fd gene II, whose products are site-specific nicking endonucleases essential for both replication of double-stranded replicative form I (RFI) to RFI replication and for the formation of single strands from RFI and RFII [28]. Unlike mutations in the ϕX174 gene A which have the unique property of being cis dominant [29], the AL1 mutations show trans complementation (data not shown) and are thus more similar to fd gene II mutations. This is not unexpected since the AL1 protein must act in trans on the TGMV B component during normal infection. Although mutations in the AL1 gene have the same effect on replication as mutations in the ϕX174 gene A or fd gene II, there is no direct evidence that the enzymatic properties of the AL1 protein are similar to those of the bacteriophage enzymes. This is the subject of further study.

The roles that the AL2 and AL3 proteins play in infection are also open to further study. Since a defect in AL2 production prevents systemic infection without affecting DNA replication, AL2 protein may be involved in virus spread. Clearly, this must be in conjunction with a protein produced from TGMV B since TGMV B functions have been previously implicated in this process[16]. Production of proteins necessary for systemic spread of TGMV from both genome components may provide a mechanism to insure that both components remain associated and available for dissemination by whitefly vectors. However, it should be noted that production of the AL2 and AL3 proteins correlates not with the insect vector but with the type of host plant. The single genome beet curly top virus[7] encodes proteins homologous to AL2 and AL3 and infects a wide range of dicotyledonous hosts, but is leaf hopper-transmitted. Single component geminiviruses that infect monocotyledonous hosts do not encode AL2 and AL3 homologues. Possibly, the AL2 ORF may encode a single-stranded DNA binding protein since the DNAs produced by these mutants seem to consistently show reduced amounts of single-stranded circular DNA forms. It is particularly interesting in this regard, that mutations in the AL2 protein do not completely abolish single-stranded DNA production as do mutants in the M13 gene V[30]. The phenotype of the AL3-1

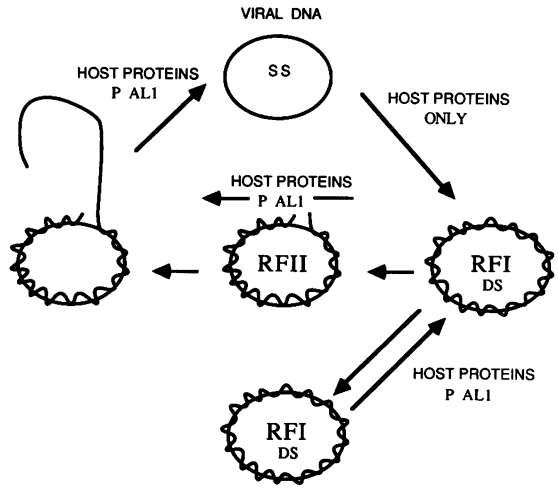

Figure 5. Model for TGMV A DNA replication. The steps in the model and evidence supporting them are described in the Discussion. SS and DS refer to single stranded and double stranded DNA, respectively. RF is replicative form. P AL1 refers to the protein product of the AL1 ORF.

mutant suggests that the AL3 protein may be involved in particle morphogenesis. This is consistent with the phenotype of delayed and attenuated symptoms which is similar to that seen for TGMV mutants defective in coat protein synthesis[14].

A scheme for geminivirus DNA replication based on ϕX174 and fd replication and our results is shown in Figure 5. An important point regarding the initial conversion of single strands to the RFI is that the primer and origin for this conversion have not been identified for any of the geminiviruses with the exception of maize streak virus[4,31]. An 80 nucleotide DNA with a few ribonucleotides at the 5' terminus is encapsidated within the virion of MSV. This oligonucleotide is located in the intergenic region between the L2 and coat protein coding sequences and acts as a primer in vitro and presumably in vivo. Similar primers have not been found associated with any of the other geminivirus DNAs even though they have been sought[6,7]. The geminiviruses may have evolved several strategies for second strand synthesis as have the single strand DNA coliphages. Although shown in this model, there is no direct evidence for RFI to RFI replication during geminivirus DNA synthesis and no mutants have been identified that only prevent formation of single-stranded circles. The model reflects the important roles played by host proteins which are required for all steps of replication. Indeed, our results indicate that the AL1 protein is

the only geminivirus protein necessary for viral DNA replication.

The results with the largest deletion described here, combined with the coat protein deletion mutant described previously[14], permit us to assign a limit to the location of the single- and double-stranded origins of TGMV DNA replication. Since deletion of the sequences located between nucleotide 1123 and nucleotide 1530 or between nucleotide 311 and 1037 both yield TGMV A DNAs that replicate and give rise to single- and double-stranded DNA, the origins of replication must be located in the remaining sequences. These are the AL1 coding sequence, a sequence of 88 nucleotides surrounding the polyadenylation signals for both the coat protein and the leftward transcripts, and the common region. These results, in combination with the fact that the common region is the only region of shared homology with the TGMV B DNA provide further evidence that the common region is the likely location of both origins of replication. Experiments to more narrowly define the locations of these replication origins are in progress. It should also be mentioned that all geminiviruses contain a sequence located in the common regions of the split genome subgroup and the large intergenic regions of the single genome subgroup which shows remarkable homology to the coliphage ϕX174 gene A protein cleavage site[16].

The results from the deletion derivatives are of practical significance in designing and constructing vectors or "plant plasmids" using geminivirus DNA. The minimal fragment capable of autonomous replication, as defined by these results, is approximately 1460 base pairs or less than 60 percent of the TGMV A sequence. Experiments are currently in progress to determine whether this TGMV replicon is maintained in transformed cells and the maximum size of foreign DNAs that may be stably replicated when inserted into this replicon.

ACKNOWLEDGEMENTS

The authors thank D. Conners and K. Kavka for synthesis of the DNA mutagenesis primers and Dr. L. Hanley-Bowdoin for reviewing and suggesting improvements to the manuscript. D.M.B., G.S. and W.E.G. were supported in part by USDA grant no. 87-CRCR-1254, NSF grant DMB-8796326 and by a grant from the Monsanto Company.

* To whom correspondence should be addressed.

REFERENCES

1. Davies, J.W., Townsend, R., Stanley, J. (1987) In Hohn, T. and Schell, J.,(eds) Plant Gene Research, Plant DNA Infectious Agents,Springer-Verlag, Vienna, Vol IV, pp. 31-52.
2. Harrison, B.D. (1985) Ann. Rev. Phytopathol. **23**, 55-82.
3. Lazarowitz, S. G. (1987) Plant Mol. Biol. Reporter **4**, 177-192.

4. Howell, S. H. (1984) Nucl. Acids Res. **12**, 7359-7375.
5. Mullineaux, P. M., Donson, J., Morris-Krsinich, B. A. M., Boulton, M. I. and Davies, J. W. (1984) EMBO J. **3**, 3063-3068.
6. MacDowell, S. W., Macdonald, H., Hamilton, W. D. O., Coutts, R. H. A. and Buck, K. W. (1985) EMBO J. **4**, 2173-2180.
7. Stanley, J., Markham, P. G., Callis, R. J., and Pinner, M. S. (1986) EMBO J. **5**, 1761-1767.
8. Stanley, J. (1983) Nature **305**, 643-645.
9. Hamilton, W. D. O., Bisaro, D. M., Coutts, R. H. A. and Buck, K. W. (1983) Nucl. Acids Res. **11**, 7387-7396.
10. Stanley, J. and Gay, M. (1983) Nature **301**, 260-262.
11. Hamilton, W. D. O., Stein, V. E., Coutts, R. H. A. and Buck, K. W. (1984) EMBO J. **3**, 2197-2205.
12. Townsend, R., Stanley, J., Curson, S. J. and Short, M. N. (1985) EMBO J. **4**, 33-37.
13. Stanley, J. and Townsend, R. (1986) Nucl. Acids Res. **14**, 5981-5998.
14. Gardiner, W. E., Sunter, G., Brand, L., Elmer, J. S., Rogers, S. G. and Bisaro, D.M. (1988) EMBO Journal **7**, 899-904.
15. Townsend, R., Watts, J. and Stanley, J. (1986) Nucl. Acids Res. **14**, 1253-1265.
16. Rogers, S.G., Bisaro, D.M., Horsch, R.B., Fraley, R.T., Hoffmann, N.L., Brand, L., Elmer, J.S. and Lloyd, A.M. (1986) Cell **45**, 593-600.
17. Messing, J., Crea, R., and Seeburg, P. H. (1981) Nucleic Acids Res. **9**, 309-321.
18. Chen, E. Y. and Seeburg, P. (1985) DNA **4**, 165-170
19. Zagursky, R., Baumeister, K., Lomax, N. and Berman, M. (1985) Gene Anal. Tech. **2**, 89-94.
20. Elmer, J. S., Brand, L., Sunter, G., Gardiner, W., Browning, C.K., Bisaro, D.M. and Rogers, S.G. (1988) Plant Molecular Biology **10**, 225-234.
21. Horsch, R.B. and Klee, H.J. (1986) Proc. Natl. Acad. Sci. USA **83**, 4428-4432.
22. Kunkel, T. A. (1985) Proc. Natl. Acad. Sci. USA **82**, 482-492.
23. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene **33**, 103-119.
24. Bale, A., d'Alarcao, M. and Marinus, M. G. (1979) Mutation Res. **59**, 157-165.
25. Dellaporta, S. L., Wood, J. and Hicks, J. B. (1983) Plant Mol. Biol. Reporter **1**, No. 4, 19-21.
26. Howarth, A. J., Caton, C., Bossert, M. and Goodman, R. M. (1985) Proc. Natl. Acad. Sci., USA **82**, 3572-3576.
27. Mullineaux, P. M., Donson, J., Stanley, J., Boulton, M. I., Morris-Krsinich, B. A. M., Markham, P. G., and Davies, J. (1985) Plant Mol. Biol. **5**, 125-131.
28. Franke, B. F. and Ray, D. S. (1971) J. Mol. Biol. **61**, 565-586.
29. Levine, A. J. and Sinsheimer, R. L. (1969) J. Mol. Biol. **39**, 619-639.
30. Pratt, D., Laws, P., and Griffith, J. (1974) J. Mol. Biol. **82**, 425-440.
31. Donson, J., Morris-Krsinich, B. A. M., Mullineaux, P. M., Boulton, M. I., and Davies, J. W. (1984) EMBO J. **3**, 3069-3073.
32. Roberts, R. J. (1987) Nucl. Acids Res., Supplement to Vol. **15**, r189-r217.
33. Brough, C. L., Hayes, R. J., Morgan, A. J., Coutts, R. H. A., and Buck, K. W. (1988) J. Gen. Virol. **69**, 503-514.