Horseradish curly top virus is a distinct subgroup II geminivirus species with rep and C4 genes derived from a subgroup III ancestor

Kimberly A. Klute, Steven A. Nadler and Drake C. Stenger

Department of Biological Sciences and Plant Molecular Biology Center, Northern Illinois University, DeKalb, IL 60115, USA

The complete nucleotide sequence (3080 nt) of an infectious DNA clone derived from the geminivirus horseradish curly top virus (HrCTV) has been determined. The relationship of HrCTV to other geminiviruses was examined using dot matrix plots of nucleotide sequence similarities, and by phylogeny of predicted amino acid sequences of individual ORFs based upon parsimony or neighbour-joining methods. These analyses indicate that the V1 and V2 virion sense ORFs of HrCTV are most closely related to, yet distinct from, the corresponding ORFs of the subgroup II geminivirus beet curly top virus (BCTV). HrCTV also encodes a third virion sense ORF (V3) which is similar (72–74% amino acid identity) to the BCTV V3 ORF; however, the HrCTV V3 ORF has diverged in sequence to a greater extent relative to that observed among isolates of BCTV (98–100% amino acid identity). The HrCTV genome encodes only three complementary sense ORFs (C1, C2 and C4) and lacks a C3 ORF which is conserved among all other subgroup II and III geminiviruses characterized to date. Although the neighbour-joining analysis indicated that the HrCTV C2 ORF was distantly related to the C2 ORF of BCTV, the predicted amino acid sequence deduced from the HrCTV C2 ORF lacks the characteristic zinc-finger domain present in the transcriptional activating protein (TrAP) encoded by the subgroup III ORF AC2, which is also retained within the TrAP-related product of the BCTV C2 ORF. Surprisingly, the rep and C4 proteins encoded by HrCTV share a closer phylogenetic relationship to the corresponding proteins of the subgroup III geminivirus squash leaf curl virus (SLCV) than to BCTV. These results suggest that the HrCTV genome may have arisen by a recombination event between a BCTV-like subgroup II virus ancestor and an SLCV-like subgroup III virus ancestor. Possible mechanisms that may explain recombination events among geminiviruses are discussed.

Introduction

The Geminiviridae is a diverse collection of plant-infecting agents characterized by a unique bisegmented virion encapsidating single-stranded circular DNA of one polarity. The family contains three recognized genera (subgroups I, II and III; Briddon & Markham, 1995) that may be distinguished based upon differences in insect vector specificity, genome organization and host range. Subgroup I includes leafhopper-transmitted viruses that have monopartite genomes and are, in all known cases except tobacco yellow dwarf virus (TYDV; Morris et al., 1992), restricted to monocot hosts. Subgroup III contains numerous whitefly-transmitted viruses infecting dicots which, with the exception of Mediterranean isolates of tomato yellow leaf curl virus (TYLCV; Navot et al., 1991) and ageratum yellow vein virus (Tan et al., 1995), possess a bipartite genome. Subgroup III viruses with bipartite genomes include African cassava mosaic virus (ACMV; Stanley & Gay, 1983), tomato golden mosaic virus (TGMV; Hamilton et al., 1984) and squash leaf curl virus (SLCV; Lazarowitz & Lazdins, 1991). Subgroup II viruses share certain superficial similarities with viruses of subgroup I, in that viruses of both subgroups are leafhopper-transmitted and have monopartite genomes. However, subgroup II viruses have a different genome organization and are found exclusively in dicot hosts. Although beet curly top virus (BCTV) is presently the only characterized member of subgroup II (Stanley et al., 1986;
Stenger et al., 1990), a preliminary report of the nucleotide sequence of tomato pseudo-curly top virus (TPCTV; Briddon et al., 1994) indicates that TPCTV also may be considered as a subgroup II virus species.

Rybicki (1994) and Padidam et al. (1995) have examined the evolutionary history of the Geminiviridae. Both of these studies provide support for the designation of three taxa within the Geminiviridae; however, subgroup II was represented in these studies by only one taxon (BCTV-California). Analyses of the cloned genomes of BCTV derived from four laboratory maintained isolates (California, Logan, CFH and Worland) have revealed considerable variation in both genotype and phenotype (Stanley et al., 1986; Lee et al., 1994; Stenger et al., 1990, 1994; Choi & Stenger, 1995), indicating that BCTV exists as an assemblage of distinct strains. Currently, the complete nucleotide sequences of the California (Stanley et al., 1986), Logan (S. G. Hormuzdi & D. M. Bisaro, unpublished, as referenced in Hormuzdi & Bisaro, 1993) and CFH (Stenger, 1994) isolates have been determined. All of these cloned isolates of BCTV retain the ability to infect sugarbeet, and typically have an extremely wide host range among dicots. In contrast, a geminivirus isolated from horseradish, with a narrow host range not including sugarbeet, was initially characterized as a strain of BCTV based upon serological relationships of the respective capsid proteins, and transmission by the same leafhopper vector species (Duffus et al., 1982).

Additional evidence supporting a relationship between BCTV and the horseradish geminivirus (hereafter referred to as horseradish curly top virus, HrCTV) was established by analysis of an infectious clone of the monopartite HrCTV genome, which bears a limited and distant similarity with BCTV based upon nucleic acid hybridization (Stenger et al., 1990). However, the previous analyses were insufficient to determine the precise relationship of HrCTV relative to BCTV and other geminiviruses. We now report the complete nucleotide sequence of HrCTV, and from analyses of sequence similarities and phylogeny obtained for various regions of the viral genome, show that HrCTV is a distinct subgroup II virus species that may have arisen by a recombination event between BCTV-like and subgroup III virus ancestors.

Methods

- **Nucleotide sequencing of HrCTV DNA.** An infectious clone of the HrCTV genome (pHrCT) has been described (Stenger et al., 1990). Subclones were constructed using the Erase-a-Base procedure (Promega), such that a nested series of deletion derivatives was obtained in both directions. DNA sequencing of the subclones was accomplished by a modification of the dideoxy chain termination method of Sanger et al. (1977) utilizing Sequenase 2.0 (United States Biochemical) and 32P-labelled dATP. The complete sequence was obtained for both strands and compiled using the DNA Inspector Ile program (Textco Inc.).

- **Sources of DNA sequences compared.** The HrCTV DNA sequence (GenBank accession no. U49907) was compared with other geminivirus sequences from each of the three subgroups. Accession numbers are listed for those sequences deposited in GenBank. Subgroup II sequences were derived from the BCTV isolates California (Stanley et al., 1986, X04144), CFH (Stenger, 1994; U02311) and Logan (S. G. Hormuzdi & D. M. Bisaro, unpublished, as referenced in Hormuzdi & Bisaro, 1993). The sequence of a dicot-infecting subgroup I virus (TYDV, M81103) was determined by Morris et al. (1992). Four subgroup III viruses were included in the analysis; TGMV (Hamilton et al., 1984; K60209), SLCV-E (Lazarowitz & Lazdins, 1991; M38182), ACMV-Kenya (Stanley & Gay, 1983; J02057) and TYLCV-Israel (Navot et al., 1991; X15656).

- **Similarity and phylogenetic analyses.** The similarity of the complete HrCTV DNA sequence to the sequences of other geminiviruses was compared using the dot matrix function of the DNA Inspector Ile program. Dot matrix parameters were defined using a window size of 20 nucleotides with a maximum mismatch tolerance of three nucleotides per window. Multiple alignments of the predicted amino acid sequences for individual geminivirus ORFs were performed with the CLUSTAL V program (Higgins et al., 1992). The TYDV rep protein amino acid sequence was translated prior to alignment based upon the deduced sequence of mature mRNA predicted to result from splicing of the C1 and C2 ORFs (Morris et al., 1992). Pairwise alignment parameters used in CLUSTAL V were gap penalty five, K-tuple one, top diagonals five and window size five. Multiple alignment parameters were kept at default values. Parsimony analyses of amino acid sequence data were completed using the PAUP program (version 3.0; Swofford, 1989), with inferred gaps treated as missing data. The branch-and-bound option of PAUP was used to find and retain all most parsimonious trees for standard and bootstrap-resampled datasets, and bootstrap percentages of clades were based on 1000 replications. Programs in the PHYLP package (version 3.57c; Felsenstein, 1989) were used to compute a distance matrix from amino acid sequences (PROTDIST, Dayhoff PAM matrix) and infer neighbour-joining trees (Saitou & Nei, 1987); bootstrap percentages in the distance trees were based on 100 replications. Trees inferred for the C1 and V1 ORFs were rooted by the subgroup I virus TYDV. The C2, C4 and V2 trees were rooted by the subgroup III virus TYLCV-I because TYDV lacks these ORFs. Neighbour-joining and parsimony analyses of the V3 ORF were not reported due to the limited number of taxa available, and the lack of an appropriate outgroup. Instead, ORF V3 amino acid sequences of BCTV isolates and HrCTV were compared simply on the basis of percentage identity.

Results

HrCTV genome organization

The complete nucleotide sequence of HrCTV DNA was determined to be 3080 nucleotides (Fig. 1); a physical map of the genome is shown in Fig. 2. The arrangement and occurrence of virion sense ORFs (V1, V2 and V3) of HrCTV are similar to those of BCTV, which suggests that the virion sense ORFs of HrCTV may serve similar functions to their counterparts in BCTV. For example, in BCTV ORF V1 encodes the capsid protein (Briddon et al., 1989). ORF V2 serves a role in the regulation of double-stranded/single-stranded DNA accumulation (Stanley et al., 1992; Hormuzdi & Bisaro, 1993) and ORF V3 encodes a protein involved in systemic movement (Hormuzdi & Bisaro, 1993; Frischmuth et al., 1993). In contrast, the organization of the complementary sense ORFs of HrCTV is somewhat different from other geminiviruses. HrCTV
apparently encodes only three complementary sense ORFs, in comparison to four ORFs present in BCTV and many subgroup III viruses. Like other geminiviruses, HrCTV possesses a rep protein encoded by C1, the largest complementary sense ORF. Based upon the size of the predicted protein encoded, the absence of obvious splicing signals and comparisons with
other geminivirus rep proteins (see below), the HrCTV C1 ORF may be considered more similar in organization to the corresponding ORF of subgroups II and III, rather than subgroup I. HrCTV also bears, in a separate reading frame, a smaller complementary sense ORF (C4) internal to the C1 ORF. The C4 ORF is also present in a similar genomic location in BCTV and many subgroup III viruses. HrCTV represents the only known subgroup II or III geminivirus to lack a C3 ORF. The C3 (or AC3) ORF of other geminiviruses encodes a protein that serves as a replication enhancer (Sunter et al., 1990; Etessami et al., 1991; Hormuzdi & Bisaro, 1995), although its mode of action has yet to be determined. HrCTV also bears an ORF (C2) located in a position in the genome similar to the AC2 ORF encoding the transcriptional activating protein (TrAP) of subgroup III (Sunter & Bisaro, 1991, 1992; Sunter et al., 1994) or the TrAP-related protein encoded by the BCTV C2 ORF (Hormuzdi & Bisaro, 1995). However, the deduced amino acid sequence of the HrCTV C2 ORF does not contain the putative zinc-finger domain characteristic of TrAP or TrAP-related proteins (Fig. 3). Furthermore, the HrCTV C2 ORF does not appear similar to the C2 ORF of subgroup I viruses, which encodes the carboxy-terminal portion of the rep protein upon production of the mature rep protein mRNA after a splicing event (Schalk et al., 1989; Morris et al., 1992).

As with all other geminiviruses characterized thus far, the HrCTV genome has an intergenic region in which cis-elements of the DNA replication origin (ori) are located (Figs 1 and 2). A potential stem–loop structure is present within the ori (nucleotides 3060–13) and bears the invariant element, TAATATTAC, containing the rep protein nick site (Laufs et al., 1995). Also present are the iterative elements described by Arguello-Astorga et al. (1994). Two of these elements (TGGAGT) are present as direct repeats (nucleotides 3014–3019 and 3022–3027) and may serve as the rep protein binding site, as is the case for other geminiviruses (Fontes et al., 1992, 1994a, b; Eagle et al., 1994). A single copy of this same sequence element also occurs at position 1952. An additional inverted copy of the iterative element is present in a location (nucleotides 2985–2990) similar to those occurring in some other geminiviruses (Arguello-Astorga et al., 1994), although a specific function has yet to be assigned for this copy. The putative C1 ORF TATA box (nucleotides 3031–3037) is located immediately downstream (in the virion sense) of the 3’-proximal directly repeated element. The spacing between the stem–loop and 3’-proximal directly repeated element of HrCTV is noticeably greater than that within the three sequenced isolates of BCTV, and more closely resembles the spacing observed in subgroup III oris. The intergenic region downstream of the ori stem–loop also contains the A–T rich tracts, which have been suggested to be involved in tertiary structure that may bend geminivirus DNA (Suarez-Lopez et al., 1995).

**Sequence comparisons and phylogenetic analyses**

Regions of extensive sequence similarity between different geminivirus genomes may be visualized in two-way comparisons when dot matrix plots are examined. When the nucleotide sequence of HrCTV was compared to other geminiviruses in dot matrix plots (Fig. 4), the virion sense genes of HrCTV were clearly similar to those of the BCTV isolates. However, the HrCTV complementary sense genes
The phylogenetic relationships of the amino acid sequences of HrCTV are consistent with the virus being a recombinant geminivirus. The HrCTV genome shares little similarity with the complementary sense genes of BCTV, with only two short regions within the C1 ORF displaying similarity under the stringent parameters used for the dot matrix analyses. Furthermore, the HrCTV genome bears little sequence similarity to the subgroup III viruses TYLCV, ACMV, and TGMV, and the similarity detected was restricted to a small region of the C1 (AC1) ORF. Surprisingly, the HrCTV C1 ORF was most similar to the AC1 ORF of the subgroup III virus SLCV (Fig. 4). The same relationship was found for the smaller C4 ORF, which entirely overlaps the larger C1 ORF. When HrCTV was compared in a dot matrix plot to the subgroup I virus TYDV, sequence similarity was detected only for the short region of the ori stem–loop (results not shown), indicating that HrCTV is not similar to a subgroup I virus. As the BCTV-California and BCTV-Logan DNA sequences are very similar (> 95%), and yielded essentially identical dot matrix plots when compared to HrCTV, only the dot matrix plot comparing BCTV-California and HrCTV is shown in Fig. 4.

The phylogenetic relationships of the amino acid sequences...
Fig. 5. Neighbour-joining trees depicting topology and distance relationships of HrCTV ORFs with other geminivirus ORFs. Trees were inferred using programs in the PHYLIP package (e.g. PROTDIST and NEIGHBOR). Horizontal branch lengths are scaled to the expected underlying number of amino acid substitutions per site. Bootstrap percentages of clades are shown (when $> 50\%$) along internal branches of the trees. Topologies of the bootstrap majority-rule consensus trees were identical to those inferred from the non-resampled datasets. Asterisks denote sequences used as outgroups to root each tree. The subgroup I virus TYDV was used preferentially as an outgroup in cases where TYDV possessed the corresponding ORF; in all other cases the subgroup III virus TYLCV was used as an outgroup. ORF V3 was not analysed due to the limited number of examples available and the lack of an appropriate outgroup. ORF C3 (AC3) was not compared because this ORF is absent in HrCTV.

deduced from individual HrCTV ORFs relative to those of other geminiviruses based upon genetic distances using PHYLIP (Fig. 5), or parsimony using PAUP (Fig. 6), are presented. Although the two methods utilize different assumptions in analysing phylogenetic relationships, the topologies of the trees inferred by the two methods were identical in all cases, except for some rearrangement of taxa in the C4 trees. Branch lengths inferred by the neighbour-joining analysis of amino acid distances (Fig. 5) show substantial differences among ORFs, suggesting varying rates of non-synonymous substitutions among the proteins. Also, within ORFs, certain clades showed long internal [e.g. C4 (SLCV, HrCTV)] or external [e.g. C2 (HrCTV)] branches (Fig. 5). Support for clades based upon bootstrap percentages was very high, particularly in the parsimony analysis (Fig. 6), suggesting that monophyletic groups were strongly supported.

From the phylogenetic analyses depicted in Figs 5 and 6, it may be discerned that the HrCTV V1, V2 and C2 ORF gene products are most closely related to, yet distinct from, the corresponding proteins of three BCTV isolates, although the HrCTV C2 ORF has clearly diverged to a greater extent. In contrast, the rep and C4 proteins of HrCTV do not share a most recent common ancestor with any of the BCTV isolates examined. Instead, the HrCTV rep protein and C4 protein were most closely related to the corresponding proteins of the subgroup III virus, SLCV. This unexpected relationship between the rep and C4 proteins of two geminiviruses classified in separate subgroups was strongly supported based upon the high percentage (parsimony, 97\% and 100\%, respectively; neighbour-joining, 100\% and 99\%, respectively) of trees examined by bootstrapping in which the two sequences were monophyletic. It is also interesting to note that for the rep protein, the branch supporting the HrCTV–SLCV clade is basal among the ingroup taxa, suggesting a relatively early divergence for this clade.

No attempt was made to estimate phylogenetic trees for the V3 ORF. This was due to the limited number of taxa available and the lack of an appropriate outgroup, as only strains of BCTV and HrCTV possess this ORF. While subgroup I viruses such as TYDV (Morris et al., 1992) encode an ORF
Table 1. Similarity of protein sequences deduced from the V3 ORF of subgroup II viruses

Presented is a matrix of percentage identity values for three BCTV isolates (California, Logan and CFH) and HrCTV in which only identical amino acid residues were considered similar.

<table>
<thead>
<tr>
<th></th>
<th>California</th>
<th>Logan</th>
<th>CFH</th>
<th>HrCTV</th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Logan</td>
<td>100</td>
<td>100</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CFH</td>
<td>98</td>
<td>98</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>HrCTV</td>
<td>72</td>
<td>72</td>
<td>74</td>
<td>100</td>
</tr>
</tbody>
</table>

(V1) present in a similar position in the genome as the subgroup II V3 ORF, there is as yet no basis for considering these ORFs to be homologous, even though both have been implicated in systemic movement. Instead, comparisons of the subgroup II V3 ORFs were restricted to reporting percentage similarities of amino acid residues (Table 1). From this limited analysis, it may be discerned that the protein encoded by the BCTV V3 ORF is highly conserved among the three strains of BCTV examined (98–100% identity). In contrast, the percentage similarity of the HrCTV V3 protein ranged between 72–74% relative to the three isolates of BCTV compared.

Discussion
Does the HrCTV sequence represent a defective genome?

The lack of a C3 ORF encoding a replication enhancer protein suggests that the characterized HrCTV genome may be debilitated. Although geminivirus genomes bearing null mutations in ORF C3 (AC3) retain infectivity, the level of DNA replication is generally greatly reduced (Sunter et al.,...
Phenotypic properties associated with HrCTV, including a limited host range (Duffus et al., 1982; Stenger et al., 1990) and low levels of viral DNA accumulation in plants inoculated with an infectious clone of HrCTV (unpublished data) may be considered consistent with a genome impaired by the lack of a functional C3 ORF. However, it remains possible that other HrCTV genotypes bearing a functional C3 ORF may have been present in the original virus culture. Certainly, the HrCTV culture used to clone the viral genome examined here was not homogeneous, as two other types of HrCTV clone were obtained during the same experiment. However, partial nucleotide sequencing of both of these clones suggests that each is defective and derived from the genome represented by pHRCT (K. A. Klute & D. C. Stenger, unpublished). One clone contained a 1.9-kbp SalI insert and resembles a defective-interfering (DI) DNA in which most of the C1 ORF and the 5′-proximal portion of the C2 ORF have been deleted. The deletion boundaries were demarcated by a three-nucleotide direct repeat, similar to the deletion boundaries of BCTV (Stenger et al., 1992; Frischmuth & Stanley, 1992) and ACMV (Stanley & Townsend, 1985) DI-DNAs. The other type of clone recovered contained a 3.0-kbp SalI insert and represents a novel class of defective genome. In this case, a portion of the 3′-proximal end of the C2 ORF was deleted and replaced by an inverted repeat derived from the adjacent region of DNA containing the 5′-proximal portion of the C2 ORF. Whether additional genotypes of HrCTV were present in the virus culture is unknown, and the virus culture from which pHRCT was derived has not been maintained. Addressing questions raised concerning variability among HrCTV genotypes will require isolation and examination of additional HrCTV cultures.

HrCTV is a recombinant geminivirus

Comparisons of the HrCTV genome with other geminivirus genomes strongly suggest that HrCTV arose through a recombination event involving a BCTV-like ancestor contributing the virion sense genes and the C2 ORF, and an SLCV-like subgroup III virus ancestor contributing the rep and C4 genes. This conclusion is supported by comparative analyses based on either similarity (distances) or parsimony. However, as it has been suggested that both SLCV and BCTV may also be recombinant geminiviruses (Torres-Pacheco et al., 1993; Rybicki, 1994), reconstruction of the specific ancestry of HrCTV, BCTV and SLCV becomes problematic. Nonetheless, HrCTV represents the best example of a recombinant geminivirus thus far reported, in which extant viruses such as BCTV and SLCV clearly possess genes that were present on the ancestral genomes from which HrCTV was derived.

Both homologous (Briddon et al., 1989; Evans & Jeske, 1993; Roberts & Stanley, 1994) and illegitimate (Stanley & Townsend, 1986) recombination have been experimentally observed with geminivirus genomes under laboratory conditions that favour selection of recombinant progeny, and either mechanism could potentially create a chimeric, twice-unit-length intermediate. In either case, recombinant progeny would result from a second crossover event, generating two reciprocal, chimeric genomes. Alternatively, the twice-unit-length intermediate could be resolved through a replicational release mechanism that has been demonstrated to be highly efficient at resolving artificially created tandemly repeated chimeras of BCTV (Stenger et al., 1991). Regardless of the recombination mechanism, the resulting progeny genomes must retain compatible cis- (Lazarowitz et al., 1992; Fontes et al., 1992; Eagle et al., 1994; Choi & Stenger, 1995) and trans-(Choi & Stenger, 1995; Jupin et al., 1995) replication elements to be potentially viable. Further limitations would be imposed by potential differences in gene regulation governed by TrAP, the necessity for trans-replication of a cognate B component, and the different genomic location of movement genes among geminivirus subgroups. In the case of a recombinant derived from subgroup II and III ancestors, only progeny resembling HrCTV (virion sense genes from subgroup II, rep gene from subgroup III) would be potentially viable, as the reciprocal recombinant would lack the ability to trans-replicate the cognate B DNA encoding movement functions required for infectivity. The same limitations would hold for potential recombinants among subgroups I and III. However, if a recombination event were to occur between viruses of subgroups I and II, the results would be less predictable, as both are monopartite and contain movement genes linked to the same DNA encoding the rep gene.

Although it cannot be stated with certainty that the HrCTV genome characterized in this report is representative of all HrCTV genotypes present in nature, the infectivity and reproducibility of the HrCTV phenotype (Stenger et al., 1990) demonstrate that this characterized clone is viable. These analyses of sequence data indicate that the HrCTV genome has a complex ancestry, which may be best explained through a recombination event which occurred sometime in the past. Given that geminiviruses have well-established means available to mediate recombination, genomes such as HrCTV should not be totally unexpected. The inference that geminiviruses are able to capture genes from other sources must be considered relevant when considering geminivirus biology and evolutionary history.

We thank E. P. Rybicki for helpful comments, and S. G. Hormuzdi and D. M. Bisaro for providing the BCTV-Logan sequence. This material is based upon work supported by the Cooperative State Research Service, U.S. Department of Agriculture, under agreement 93-37303-9384 awarded to D.C.S.

References


Received 19 December 1995; Accepted 6 March 1996