For the Record

Complete Nucleotide Sequence of the Hypervirulent CFH Strain of Beet Curly Top Virus

Drake C. Stenger

Department of Biological Sciences, Northern Illinois University, DeKalb 60115, U.S.A.
Received 29 July 1993. Accepted 25 October 1993.

The complete nucleotide sequence of the hypervirulent CFH strain of beet curly top geminivirus (BCTV) has been determined. The circular DNA genome of BCTV-CFH consists of 2,927 nucleotides and shares extensive sequence homology with the biologically distinct California strain of BCTV. Analysis of the CFH nucleotide sequence indicated that the rightward open reading frames (ORFs) R1, R2, and R3 are highly conserved (>95% amino acid chemical similarity) in the CFH and California strains, although CFH ORF R2 was extended by 24 carboxy-terminal amino acid residues not present in the California strain. The CFH leftward ORFs L1, L2, and L3 shared varying levels of amino acid chemical similarity with the corresponding ORFs of the California strain (78.8, 66.5, and 86.7%, respectively). CFH ORF L4 was the least conserved ORF present in both strains, encoding a 9.9-kDa protein of 87 amino acid residues, which shares 57.6% chemical similarity with 85 carboxy-terminal amino acid residues of the 19.4-kDa ORF L4 of the California strain. The CFH DNA sequence also contained a unique 12.5-kDa ORF (R4); however, there is no evidence to suggest that R4 is expressed. Comparison of the CFH and California strain nucleotide sequences indicates that certain regions of the BCTV genome have diverged, and this divergence may account for differences in the pathogenic properties of the two strains.

Additional keywords: genome organization.

Beet curly top virus (BCTV) can be distinguished from other geminiviruses on the basis of genome organization, insect transmission, and host range (for review see Lazarowitz [1987], Davies and Stanley [1989], and Stanley [1991]). Stanley *et al.* (1986) determined the complete nucleotide sequence of a typical BCTV strain (California), and by using an infectious DNA clone further demonstrated that the single DNA genomic component (2,993 nucleotides) of BCTV encodes all viral functions necessary for replication, systemic movement, induction of disease symptoms, encapsidation, and transmission by leafhoppers. The BCTV genome encodes seven open reading frames (ORFs) which appear to be functional, based upon the analysis of mutant phenotypes. Briddon *et al.* (1989) established that BCTV ORF L1 is required for viral replication, while the capsid protein ORF R1 is essential for infectivity, because of requirement for the capsid protein in systemic movement. The BCTV capsid protein is also responsible for the specificity of leafhopper transmission (Briddon *et al.* 1990). BCTV ORF R2 has been associated with the production and accumulation of viral ssDNA (Stanley *et al.* 1992; Hormuzdi and Bisaro 1993), while ORF R3 has been identified as a second BCTV gene mediating systemic movement of the virus (Hormuzdi and Bisaro 1993). Mutations introduced in ORFs L2, L3, and L4 have been associated with alterations in symptom expression (Stanley and Latham 1992; Stanley *et al.* 1992), suggesting that these ORFs are also functional, although their primary functions remain to be identified.

While numerous strains of BCTV have been isolated and characterized with respect to biological properties (Bennett 1971), little information is available concerning the molecular basis for the diversity of pathogenic properties exhibited by distinct strains. In addition to the California strain, infectious DNA clones for four other BCTV strains (Logan, CFH, Worland, and HRCT) have been isolated and characterized for pathogenic and physical properties of the genome (Stenger *et al.* 1990). The California and Logan strains are representative of typical BCTV strains with wide host ranges. Restriction endonuclease patterns for these two strains are very similar, and they are biologically indistinguishable in their pathogenic properties (Stenger *et al.* 1990). The complete nucleotide sequence of the Logan strain (3,038 nucleotides) has been determined, and the Logan strain sequence differs little from that of the California strain sequence (S. G. Hormuzdi and D. M. Bisaro, unpublished data, cited in Hormuzdi and Bisaro 1993). In contrast, the hypervirulent CFH strain of BCTV can be distinguished from the California and Logan strains by a distinct endonuclease restriction map and also by causing severe symptoms on *Nicotiana tabacum* L., which remains asymptomatic when infected by the Logan or the California strain (Stenger *et al.* 1990). Current experiments indicate that the CFH strain can be further distinguished from other BCTV strains by producing more severe symptoms in *N. benthamiana* Domin and *Arabidopsis thaliana* (L.) Heynh. (D. C. Stenger, D. M. Bisaro, and K. R. Davis, unpublished data). To characterize the molecular basis of BCTV strain diversity, the nucleotide sequences of bio-

Corresponding author: Drake C. Stenger. Telephone 815-753-7805. Fax 815-753-0461 U.S.A.

Nucleotide sequence data have been submitted to GenBank as accession number U02311.

MPMI Vol. 7, No. 1, 1994, pp. 154-157

logically distinct variants of the virus need to be determined. Regions of the viral genome potentially responsible for the observed differences between strains may then be identified for further analysis. In this communication, I report the complete nucleotide sequence of the hypervirulent CFH strain of BCTV and compare the nucleotide sequence of the CFH strain to that of the biologically distinct California strain.

pCFH represents a full-length, infectious DNA clone of the CFH strain of BCTV, which has been described by Stenger *et al.* (1990). Subclones were constructed by use of the Erase-a-Base procedure (Promega, Madison, WI) such that a nested series of deletion derivatives were obtained in both directions. DNA sequencing of subclones was accomplished by a modification of the dideoxy chain termination method of Sanger *et al.* (1977) utilizing Sequenase 2.0 (United States Biochemical, Cleveland, OH) and ^{35}S-labeled dATP. The complete nucleotide sequence of the BCTV-CFH genome was determined to consist of 2,927 nucleotides and is presented in Fig-

```
   1  ATTGAATCGGGCTCTCTTCAAATCCCCTATCAATTGGGTGCTTTGGGTGCTCTTAAATAC   60
  61  CACCAAGGGGCCATCCGCATTAATATTACCGGATGGCCCCCAAAAAATACGTGGCCCAAT  120
 121  GAAAATATGCCACGTGGAAAGCTAAAGATGTTGATGTGATGGATTTGGGCGGGAAACTTC  180
 181  CTGAAGAAGTTTTCATTCCCGCCCCTTTTTTTGACTTAGTCAAAAAAGTAGCCGACATAC  240
 241  TGTTTGAAATAAAGTTAAAAGTAGATCATTGTGGTCCCCACAAGAAACTTGCTAAGCAAG  300
 301  TTTCGAATTATTTCTGTTTAAATACGTTATTTTTACATGTATATGTACTTATAAATGATG  360
 361  GTCTGTCTACCAGACTGGTTATTTTTGCTATTTATCTTCAGTATTCTCCTGCAATCAGGT  420
 421  ACGAACTTTTATGGGACCTTTCAGAGTGGATCAATTTCCAGACAATTATCCAGCCTTTCT  480
 481  AGCAGTATCGACCAGTTGTTTCTTAAGGTACAACAAGTGGTGTATACTAGGTATCCATCA  540
 541  AGAGATAGAGGCTCTGACCCTAGAAGAAGGCGCGGTCTTTCTTCAATTCCAGAAGGAAGT  600
 601  GAAGAAGCTACTGAGGCGTAAGGTTAATTTTCATAGGAAGTGTTCGTTGTATGAGGAAAT  660
 661  ATACAAGAAATACGTATACAATGTCCCAGAAAAGAAAGGTGAATCCTCAAAGTGCGTGGC  720
 721  CGAAGAAGAGGAGGACTACTACGATTTCGAGGAAATACCAATGGAGGAGACCTGTGACAA  780
 781  AAAACAGGACTCTGAAGTTAAAGATGTATGATGATATGTTGGGTGCTGGTGGTATAGGAT  840
 841  CTACCATTAGTAATAATGGTATGATTACTATGTTGAATAATTATGTCCAGGGTATTGGTG  900
 901  ATAGTCAGAGAGCGAAGAACGTTACTGTGACGAAGCATTTGAAGTTTGATATGGCTCTTA  960
 961  TGGGGAGTTCTCCGTTCTGGGAGACTCCTAATTATATGACCCAATATCATTGGATTATCA 1020
1021  TAGACAAGGATGTTGGGTCAGTGTTTCCTACTAAGTTATCGAGTATATTTGATATTCCCG 1080
1081  ATAACGGTCAGGCTATGCCGTCTACTTATCGTATTCGAAGAGATATGAACGAGAGGTTTA 1140
1141  TTGTGAAGAAGAAATGGAGAACTCATTTGATGTCTACTGGTACTGGATATGGAGGGAAGG 1200
1201  AGACTTACAAGGCTCCTTCAATGCCAAATTACAAGAAACCGATGAATATCAATGTCCGCA 1260
1261  ATCTGAACATGAGGACTATTTGGAAAGACACCGGTGGTGGGAAGTATGAAGACGTGAAGG 1320
1321  AGAATGCTTTACTCTATGTTGTTGTTAATGATAATACGGATAATACTAATATGTATGCCA 1380
1381  CTTTGTTTGGCAATTGTAGATGTTATTTTTATTAATAAAAATATTATTTTTTTAATACAA 1440
1441  TTTCATTGCAACACTCTGGTTCTGAATTACACTAGATACAAAATTAAATTCTTCATACAG 1500
1501  GACATGATTTAAACCTCTAATTACATTATTAATACAAATTAAACCTAAATTATCCAAGTA 1560
1561  TATCTCTAGTCTCTTTTTCAAGACTTTCAAAAAACGGTCCCCAGTTAATGTCCTGTGCGA 1620
1621  TCCAGTGATCGTCAAATCTATCCAGCACTTGTGAAGATTCAAGTATTTGCGGAGGTTGTG 1680
1681  GTTGAATCTTATCTGGACTTTCAGTTGATAAACTGGCCCGGAACGGAAGAAGTCCTGGAG 1740
1741  TTTGAGGTACAATGGATTCGGTACCAAGTCCACGGGTATGGAATTCGTCGCTTGTTGCAG 1800
1801  CGTGATGGGTTCCTCTGTGCGAAAATCCCTGATTACATTCATGATGAATTGTGAAATGAC 1860
1861  ACTTACAAGGAAGTTTGATCTTGCGAGGACGCTTCTTGATTGTTATCAAAGAGAGGACTT 1920
1921  GTGAGTTTGGCGAAGACTGAATTTTGTAATGTCCAGGACCTAAGGGCTTCATTTTCTGAT 1980
1981  TTATTGAGGAAGTCCTGGTAAGAGCTGCCTTCGCCTGGATTGCATAATATAATACTGGGA 2040
2041  ATACCACCTTTAATGACACGTGGTTTTCCATACTTTAAGTTTGTCTGCCACTCTCTTTGT 2100
2101  GCGCCTATAAGGTGCTTCCAATGCTTCATCTTTAAGTAATTGGGATCTACGTCATCAATG 2160
2161  ACGTTATACAGAACATTATCATTATAAGTTTTTAAACTAAAATCTAAATGACCAGTAATA 2220
2221  TAATTATGTGGACCTAAACACCTGGCCCACATTGTTTTACCAGTTCTTGAATCACCCTCT 2280
2281  ATGATTAAACTATTATATCTAAAAGGCCGCGCAGCGGGATCCAAACCAAAATAAGAATCG 2340
2341  GCCCACTCTTGAATAATTTCTGGAACTCTAGTGAAAGAAGAAAGAGGAAAAGGTGGTTGA 2400
2401  TAAAGATCTGGAGGAGGAAGAAAAATAGCCTCTAAGTTAGGTTTGAGGTTGTGATACTGA 2460
2461  AAAATATATTTTTCTGGGAGTTTCTCCCTTATTATCTGCAGAGCTTCTAATGCATTACCT 2520
2521  GCATTTAATGCCTCTGCTGCAGCATCATTAGCCGTCTGTTGACCTCCGCGTGCAGATCTT 2580
2581  CCATCGACCTGAAATTCACCCCAGTCGATGTGATCTCCGTCCTTTGAGACGTAGGACTTG 2640
2641  ACGTCGGAACTGGATTTAGCTCCCTGAATATTGCAGTGGAATTGTTTGCTGGTACTTCGA 2700
2701  TGTTGCAGATCGAAGTAACGGGCATTACGGATCTGGACTTTTCCTTCGAATTGAATAAGG 2760
2761  GCATGCAGATGTGGTTCCCCATTTTCATGTAATTCTCTGCAGATGCGAATATATTTTTTA 2820
2821  TTCGAAGGTGTATTTATAGCGAGGAGCTGTTCTAAGGCGTCTTCTTTGGTTACTGAACAT 2880
2881  TGAGGGTATGTAAGGAAAAAATTTTTGGCTTTTTTGTAAAAAGGCAT              2927
```

Fig. 1. Nucleotide sequence of beet curly top virus strain CFH. The sequence presented is in the virion sense, with nucleotide 1 defined as the first nucleotide upstream of the ATG initiation codon of open reading frame L1.

ure 1. A physical map of the circular CFH genome (Fig. 2) depicts the locations and nucleotide coordinates of restriction endonuclease sites and the location and polarity of ORFs potentially encoding proteins of at least 85 amino acid residues. As in other geminiviruses, ORFs are present in both the viral-sense (rightward ORFs R1, R2, R3, and R4) and the complementary-sense (leftward ORFs L1, L2, L3, and L4) strands of the CFH genome. The ORF nomenclature used follows that of the Logan strain (Hormuzdi and Bisaro 1993).

Of the eight CFH ORFs, seven are conserved in the CFH and California strains (Table 1). The CFH genome also contains a unique ORF (R4), which is not present in the California strain and does not exhibit significant homology with other geminivirus ORFs. At present there is no evidence to indicate that CFH ORF R4 is expressed, and the presence of this ORF in the CFH genome may be fortuitous. An examination of the amino acid sequences of the corresponding ORFs of the CFH and California strains indicated that the rightward ORFs R1, R2, and R3 are highly conserved and retain >95% chemical similarity of amino acid sequence (Table 1), although the CFH R2 ORF was extended by 24 amino acid residues at the carboxy terminus. In contrast, the chemical similarity of amino acid sequences in the leftward ORFs varied from 86.7% for ORF L3 to as low as 57.6% for ORF

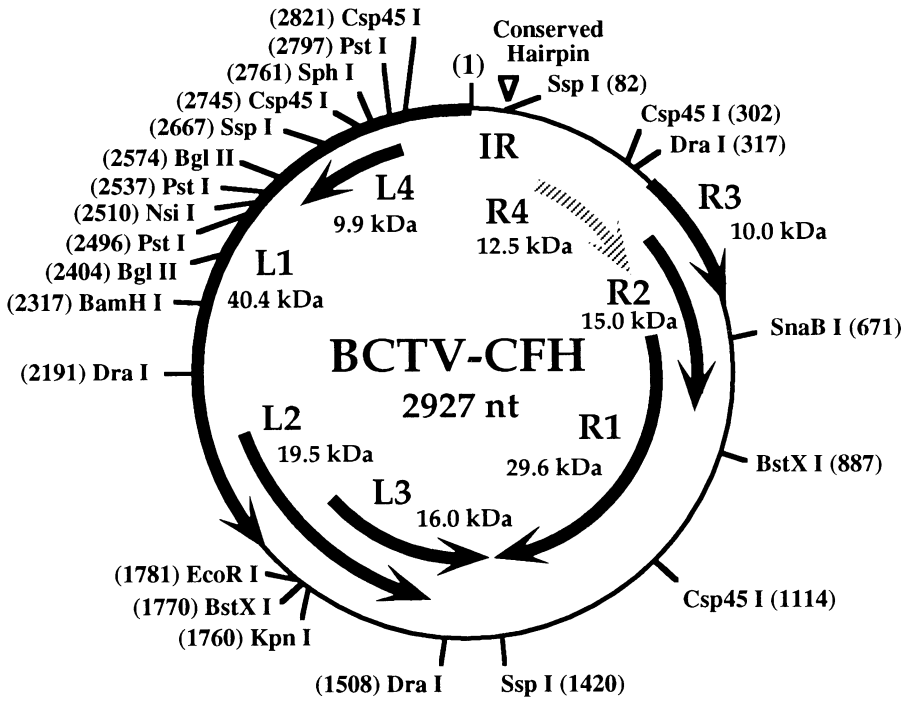

Fig. 2. Physical map of the genome of beet curly top virus (BCTV) strain CFH. The solid arrows denote the location and polarity of open reading frames (ORFs) conserved among the CFH and California strains of BCTV. The shaded arrow denotes the location and polarity of a nonconserved ORF unique to BCTV-CFH. The sizes of the predicted protein products are indicated in kilodaltons. The nucleotide coordinates of restriction endonuclease cleavage sites are indicated in parentheses. The location of the intergenic region (IR) and the conserved geminivirus hairpin sequence motif (inverted triangle) are indicated.

Table 1. Analysis of open reading frames (ORFs) of beet curly top virus (BCTV) strain CFH with the potential to code for proteins greater than 85 amino acid residues

ORF	Protein mass (kDa)	Frame	Nucleotide coordinates		Amino acids (no. of residues)	Amino acid homology (%)[a]	
			Start	Stop		Direct	Conserved
R1	29.6	+3	651	1,413	254	98.0	98.8
R2	15.0	+2	431	809	126	94.1[b]	95.1[b]
R3	10.0	+1	355	619	88	97.7	100
R4	12.5	+3	159	480	107	...[c]	...[c]
L1	40.4	−1	2,927	1,865	354	71.5[d]	78.8[d]
L2	19.5	−2	2,071	1,552	173	62.4	66.5
L3	16.0	−3	1,842	1,434	136	82.4	86.7
L4	9.9	−2	2,788	2,527	87	52.9[e]	57.6[e]

[a] Sequence homology with the California strain of BCTV. Direct homology was based upon identical amino acid residues only. Conserved homology was based upon both direct homology and chemically similar residues considered identical.
[b] Only amino-terminal residues 1–102 were compared. CFH ORF R2 contains 24 additional carboxy-terminal residues not present in BCTV-California.
[c] ORF R4 is not conserved among other BCTV strains.
[d] Homology determined after alignment of sequences to account for residues 7–10 of BCTV-California, which are not present in CFH.
[e] CFH ORF L4 residues 3–87 compared with BCTV-California ORF L4 carboxy-terminal residues 83–167.

L4 (Table 1). In two cases (ORFs L1 and L4), comparison of amino acid sequences for the predicted proteins required alignment prior to analysis. CFH ORF L1 is four amino acid residues shorter than the corresponding ORF in the California strain, and an alignment of the predicted amino acid sequences indicated that residues 7–10 of the California ORF L1 are not present in the CFH ORF L1 sequence. ORF L4 of CFH (9.9 kDa) was significantly shorter than the corresponding ORF L4 of the California strain (19.4 kDa). Although the CFH genome retained the ATG initiation codon (CFH nucleotide coordinate 74) corresponding to the California ORF L4 initiation codon, this ORF in CFH terminated after only 14 amino acid residues. However, the 9.9-kDa CFH ORF L4 was initiated at a site downstream relative to the California strain, such that CFH ORF L4 retained sequences corresponding to the carboxy-terminal amino acid residues of the California strain ORF L4. This observation is consistent with the hypothesis that the California strain ORF L4 may actually begin translation at a downstream ATG (Stanley *et al.* 1992), as suggested by mutational analyses of the California strain ORF L4 in which mutations constructed in the amino portion had no effect, while mutations constructed in the conserved carboxy portion resulted in attenuated symptoms (Stanley and Latham 1992; Stanley *et al.* 1992).

Within the CFH intergenic region (IR) is located a conserved hairpin sequence motif, NNNTAATATTAC, flanked by inverted repeats of 11 nucleotides (nucleotide coordinates 67–100). This motif is highly conserved among distinct geminiviruses (Lazarowitz 1987) as an essential element of the geminivirus plus-strand replication origin (Stenger *et al.* 1991; Lazarowitz *et al.* 1992). While the California strain IR contains a 27-nucleotide direct repeat (Stanley *et al.*, 1986), the CFH IR retains only a single copy of a similar sequence element (nucleotide coordinates 169–195), indicating that any potential function of this element does not require duplication. Aside from these similarities, the remainder of the IR of the two strains has undergone considerable divergence.

Determination of the BCTV-CFH sequence has permitted identification of regions of the BCTV genome which have diverged in different strains, some of which must specify determinants of strain-specific pathogenic properties. Completion of the CFH sequence also facilitates ongoing research concerning the production of chimeric viral genomes containing defined fragments of DNA from biologically distinct strains, in an effort to identify which regions of the BCTV genome control strain-specific traits.

ACKNOWLEDGMENTS

This material is based upon work supported by the Cooperative State Research Service, U.S. Department of Agriculture, under agreement 93-37303-9384.

LITERATURE CITED

Bennett, C. W. 1971. The Curly Top Disease of Sugarbeet and Other Plants. Monogr. 7. American Phytopathological Society, St. Paul, MN.

Briddon, R. W., Pinner, M. S., Stanley, J., and Markham, P. G. 1990. Geminivirus coat protein gene replacement alters insect specificity. Virology 177:85-94.

Briddon, R. W., Watts, J., Markham, P. G., and Stanley, J. 1989. The coat protein of beet curly top virus is essential for infectivity. Virology 172:628-633.

Davies, J. W., and Stanley, J. 1989. Geminivirus genes and vectors. Trends Genet. 5:77-81.

Hormuzdi, S. G., and Bisaro, D. M. 1993. Genetic analysis of beet curly top virus: Evidence for three virion sense genes involved in movement and regulation of single- and double-stranded DNA levels. Virology 193:900- 909.

Lazarowitz, S. G. 1987. The molecular characterization of geminiviruses. Plant Mol. Biol. Rep. 4:177-192.

Lazarowitz, S. G., Wu, L. C., Rogers, S. G., and Elmer, J. S. 1992. Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. Plant Cell 4:799-809.

Sanger, F., Nicklen, S., and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Stanley, J. 1991. The molecular determinants of geminivirus pathogenesis. Semin. Virol. 2:139-149.

Stanley, J., and Latham, J. R. 1992. A symptom variant of beet curly top geminivirus produced by mutation of open reading frame C4. Virology 190:506-509.

Stanley, J., Latham, J. R., Pinner, M. S., Bedford, I., and Markham, P. G. 1992. Mutational analysis of the monopartite geminivirus beet curly top virus. Virology 191:396-405.

Stanley, J., Markham, P. G., Callis, R. J., and Pinner, M. S. 1986. The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. EMBO J. 5:1761-1767.

Stenger, D. C., Carbonaro, D., and Duffus, J. E. 1990. Genomic characterization of phenotypic variants of beet curly top virus. J. Gen. Virol. 7:2211-2215.

Stenger, D. C., Revington, G. N, Stevenson, M. C., and Bisaro, D. M. 1991. Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of a plant viral DNA. Proc. Natl. Acad. Sci. USA 88:8029-8033.