Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene

Robert M. Harding, 1 Thomas M. Burns, 1 Gregory Hafner, 1 Ralf G. Dietzgen 2 and James L. Dale 1*

1 Centre for Molecular Biotechnology, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001 and 2 Plant Protection Division, Queensland Department of Primary Industries, Indooroopilly, Queensland, Australia

One DNA component of the banana bunchy top virus (BBTV) genome was cloned and sequenced. This component is present as a circular, ssDNA in the virions and consists of 1111 nucleotides. It contains one large open reading frame (ORF) of 858 nucleotides in the virion sense; this ORF encodes a putative replicase based on the presence of a dNTP-binding motif (GGEKG). Two smaller ORFs (249 and 366 nucleotides), in the complementary orientation, could not be assigned any obvious function. Neither of these ORFs had significant sequence homology with any known DNA plant virus gene or gene product. Computer analysis of this component predicted a strong stem–loop structure in the virion sense putative untranslated region; a nonanucleotide sequence in the loop was nearly identical to the nonanucleotide invariant loop sequence of geminiviruses and coconut foliar decay virus. There is strong evidence that the genome of BBTV consists of more than one component because no ORF was found that would encode a protein the size of the BBTV coat protein. BBTV has some characteristics in common with geminiviruses but cannot be classified as one. Rather, BBTV probably belongs to an undescribed plant virus group which could also include subterranean clover stunt virus and coconut foliar decay virus.

Introduction

Banana bunchy top disease is the most serious virus disease affecting bananas (Musa spp.). Originally it was assumed to be caused by a luteovirus on the basis of the biological characteristics of the disease: it is persistently transmitted by aphids, causes a yellow type disease and infected plants have damaged phloem (Dale, 1987). Further evidence for the involvement of a luteovirus included (i) association of dsRNA with the disease (Dale et al., 1986), (ii) purification of 28 nm isometric virus-like particles (VLPs) from infected plants (Iskra et al., 1989) and (iii) purification of 20 to 22 nm isometric VLPs containing ssRNA with an $M_r$ of 2.0 x 10^6 from infected plants (Wu & Su, 1990). However, Harding et al. (1991) and Thomas & Dietzgen (1991) purified 18 to 20 nm isometric VLPs from infected plants and these particles contained ssDNA of about 1 kb. SCSV is persistently transmitted by aphids, whereas CFDV is transmitted by a planthopper. Furthermore, the genome of SCSV is composed of at least seven distinct ssDNA molecules each containing circular ssDNA of approximately 1 kb. SCSV is persistently transmitted by aphids, whereas CFDV is transmitted by a planthopper. The sequence of one ssDNA molecule of CFDV has been determined (Rohde et al., 1990) and one of the ORFs encodes a putative replicase. There are also three animal-infecting viruses that have similar characteristics to BBTV, SCSV and CFDV, namely porcine circovirus (PCV), chicken anaemia virus (CAV) and psittacine beak and feather disease virus (PBFDV) (Todd et al., 1991).

In this paper, we report the cloning, sequencing and analysis of one component of BBTV DNA.

Methods

Sequencing. Mini-preparations of pBT338 (Harding et al., 1991) were prepared by alkaline lysis followed by polyethylene glycol precipitation (Hattori & Sakaki, 1986). Sequencing was done using
[32P]dCTP and a Sequenase kit (US Biochemicals) as recommended by the manufacturer. Reaction products were analysed on an 8\% (w/v) polyacrylamide gel containing 7 M-urea. Gels were fixed, dried and exposed to Agfa Curix RP1 film. The primers used for sequencing were either universal sequencing primers (US Biochemical) or 17 to 30 nucleotide primers complementary to appropriate regions of the cloned viral DNA (Fig. 1). The latter primers were synthesized using an Applied Biosystems 391 DNA synthesizer.

**PCR: analysis and cloning.** From the nucleotide sequence of pBT338, two oligonucleotides (primer A: 5' GGAAGAAGCCTCTCATCT-GCTTCAGAGAGC 3'; primer B: 5' CAGGCGCACACCTTGAGAAACGAAAGGGAA 3') were synthesized and were used as primers in a PCR with purified BBTV DNA. The reaction mix (50 µl) contained 10 mM-Tris–HCl pH 8.3, 50 mM-KCl, 1.5 mM-MgCl₂, 200 µM each dNTP, 50 pmol each primer and 0.6 units Taq polymerase (Cetus). Following the addition of template DNA (approximately 0.1 ng), the mix was subjected to one cycle consisting of denaturation at 94 °C for 5 min, annealing at 55 °C for 2 min and extension at 72 °C for 3 min; 30 cycles consisting of denaturation at 94 °C for 1 min, annealing at 55 °C for 2 min and extension at 72 °C for 3 min; and finally one cycle consisting of denaturation at 94 °C for 1 min, annealing at 55 °C for 2 min and extension at 72 °C for 10 min. The amplified product was analysed by electrophoresis in either 1% agarose gels in Tris-acetate–EDTA buffer, pH 7.8 (Maniatis et al., 1982) or in discontinuous polyacrylamide gels (5% stacking gel/10% resolving gel) using the buffer system of Laemmli (1970) without SDS. Nucleic acids were visualized with ethidium bromide and the size of the amplified product was estimated by comparison with a BglI/HindII digest of pBR328 (Boehringer Mannheim).

The amplified product was cloned directly into the plasmid vector, pCR2000, using a TA cloning kit (Invitrogen) as recommended by the manufacturer. Potential recombinant clones were identified by screening on X-gal substrate, and virus-specific clones were subsequently identified by screening purified plasmids with 32P-labelled insert from pBT338 (Harding et al., 1991). Plasmids that hybridized with the pBT338 insert and contained inserts of approximately 1.1 kb were sequenced both by screening purified plasmids with 32P-labelled insert from pBT338, two primers (primers A and B, each 30 nucleotides in length) were synthesized, which at the same time would demonstrate that BBTV was circular. Furthermore it was assumed that this product represented the full sequence of one DNA component of the BBTV genome for the following reasons: (i) it was the only amplification product of a PCR that should yield full-length product from a circular template and (ii) it was approximately the size of the ssDNA extracted from purified BBTV virions.

The amplified product was ligated into a ‘T-tailed’ plasmid (pCR2000) and this plasmid was transformed into Escherichia coli. A small sample of this transformation was analysed and potential recombinants were screened using pBT338 as a probe. Five potential recombinants hybridized with the pBT338 insert and contained inserts of approximately 1.1 kb. Three of these

\[
\text{Results}
\]

The insert from pBT338 (Harding et al., 1991) was completely sequenced in both directions and was found to contain 980 bp. From this, we developed a strategy for the synthesis of double-stranded full-length BBTV DNA which at the same time would demonstrate that BBTV virion ssDNA was circular. Using the sequence information obtained from pBT338, two primers (primers A and B, each 30 nucleotides in length) were synthesized, which hybridized immediately adjacent to one another but were reversed in their orientation. These primers were used in a PCR with Taq DNA polymerase. In this reaction, primer A hybridized to the virion ssDNA and cDNA was synthesized and primer B hybridized to the 3’ end of the resultant cDNA and a copy of the template DNA was synthesized. This reaction was analysed by PAGE; a single amplified product was evident with a size of about 1.1 kb which appeared to be a full-length double-stranded copy of the template DNA (Fig. 2). This result provided strong evidence that the virion ssDNA of BBTV was circular. Furthermore it was assumed that this product represented the full sequence of one DNA component of the BBTV genome for the following reasons: (i) it was the only amplification product of a PCR that should yield full-length product from a circular template and (ii) it was approximately the size of the ssDNA extracted from purified BBTV virions.

The amplified product was ligated into a ‘T-tailed’ plasmid (pCR2000) and this plasmid was transformed into Escherichia coli. A small sample of this transformation was analysed and potential recombinants were screened using pBT338 as a probe. Five potential recombinants hybridized with the pBT338 insert and contained inserts of approximately 1.1 kb. Three of these

\[
\text{Results}
\]

The insert from pBT338 (Harding et al., 1991) was completed sequenced in both directions and was found to contain 980 bp. From this, we developed a strategy for the synthesis of double-stranded full-length BBTV DNA which at the same time would demonstrate that BBTV virion ssDNA was circular. Using the sequence information obtained from pBT338, two primers (primers A and B, each 30 nucleotides in length) were synthesized, which hybridized immediately adjacent to one another but were reversed in their orientation. These primers were used in a PCR with Taq DNA polymerase. In this reaction, primer A hybridized to the virion ssDNA and cDNA was synthesized and primer B hybridized to the 3’ end of the resultant cDNA and a copy of the template DNA was synthesized. This reaction was analysed by PAGE; a single amplified product was evident with a size of about 1.1 kb which appeared to be a full-length double-stranded copy of the template DNA (Fig. 2). This result provided strong evidence that the virion ssDNA of BBTV was circular. Furthermore it was assumed that this product represented the full sequence of one DNA component of the BBTV genome for the following reasons: (i) it was the only amplification product of a PCR that should yield full-length product from a circular template and (ii) it was approximately the size of the ssDNA extracted from purified BBTV virions.

The amplified product was ligated into a ‘T-tailed’ plasmid (pCR2000) and this plasmid was transformed into Escherichia coli. A small sample of this transformation was analysed and potential recombinants were screened using pBT338 as a probe. Five potential recombinants hybridized with the pBT338 insert and contained inserts of approximately 1.1 kb. Three of these

\[
\text{Results}
\]

The insert from pBT338 (Harding et al., 1991) was completed sequenced in both directions and was found to contain 980 bp. From this, we developed a strategy for the synthesis of double-stranded full-length BBTV DNA which at the same time would demonstrate that BBTV virion ssDNA was circular. Using the sequence information obtained from pBT338, two primers (primers A and B, each 30 nucleotides in length) were synthesized, which hybridized immediately adjacent to one another but were reversed in their orientation. These primers were used in a PCR with Taq DNA polymerase. In this reaction, primer A hybridized to the virion ssDNA and cDNA was synthesized and primer B hybridized to the 3’ end of the resultant cDNA and a copy of the template DNA was synthesized. This reaction was analysed by PAGE; a single amplified product was evident with a size of about 1.1 kb which appeared to be a full-length double-stranded copy of the template DNA (Fig. 2). This result provided strong evidence that the virion ssDNA of BBTV was circular. Furthermore it was assumed that this product represented the full sequence of one DNA component of the BBTV genome for the following reasons: (i) it was the only amplification product of a PCR that should yield full-length product from a circular template and (ii) it was approximately the size of the ssDNA extracted from purified BBTV virions.

The amplified product was ligated into a ‘T-tailed’ plasmid (pCR2000) and this plasmid was transformed into Escherichia coli. A small sample of this transformation was analysed and potential recombinants were screened using pBT338 as a probe. Five potential recombinants hybridized with the pBT338 insert and contained inserts of approximately 1.1 kb. Three of these
Fig. 3. Nucleotide sequence of BBTV component 1 ssDNA. BBTV virion ssDNA hybridized with 32P end-labelled primer A (a) and primer B (b). Lanes 1, pBT338; lanes 2, pBTPCR7, -11 and -12; at nucleotide 1045, C occurred in pBT338 and pBTPCRll whereas a T occurred in pBTPCR7 and -11 was replaced by a T in pBTPCR12 resulting in the replacement of an asparagine residue by arginine residue; at nucleotide 508, the A in pBT338, pBTPCR7 and -11 was replaced by a T in pBTPCR12 resulting in the replacement of a histidine residue by an arginine residue; at nucleotide 701, A occurred in pBTPCR7 and -11 was replaced by a T in pBTPCR12, resulting in the replacement of an asparagine residue by arginine residue; at nucleotide 256, the A in pBT338, pBTPCR7 and -11 was replaced by a G in pBTPCR12 underlined; the potential TATA box is in bold and boxed; the potential stem-loop sequence is in bold and double-underlined; and the putative dNTP-binding motif (GGEGKT) is in bold italics and underlined.

To determine which sequence orientation of the BBTV DNA component 1 ssDNA was present in virions, BBTV ssDNA was extracted from purified virions, electrophoresed through agarose and transferred to nylon membranes. These membranes were incubated with one of two 32P end-labelled oligonucleotides (primer A or B). It was found that primer A hybridized with virion ssDNA (Fig. 4a) whereas primer B did not (Fig. 4b). Primer A was complementary to the sequence presented in Fig. 3 confirming that this was the orientation present in virions.

Fig. 4. Southern blot analysis to determine the orientation of BBTV component 1 ssDNA. BBTV virion ssDNA hybridized with 32P end-labelled primer A (a) and primer B (b). Lanes 1, pBT338; lanes 2, pBTPCR7, -11 and -12; at nucleotide 1045, C occurred in pBT338 and pBTPCRll whereas a T occurred in pBTPCR7 and -11 was replaced by a T in pBTPCR12 resulting in the replacement of an asparagine residue by arginine residue; at nucleotide 508, the A in pBT338, pBTPCR7 and -11 was replaced by a T in pBTPCR12 resulting in the replacement of a histidine residue by an arginine residue; at nucleotide 701, A occurred in pBTPCR7 and -11 was replaced by a T in pBTPCR12, resulting in the replacement of an asparagine residue by arginine residue; at nucleotide 256, the A in pBT338, pBTPCR7 and -11 was replaced by a G in pBTPCR12 underlined; the potential TATA box is in bold and boxed; the potential stem-loop sequence is in bold and double-underlined; and the putative dNTP-binding motif (GGEGKT) is in bold italics and underlined.

polymerases used to generate the clones, or reflected genuine sequence variations in the viral genome. In most instances, the sequence obtained from pBT338 was the most common and was used to derive the final sequence for BBTV DNA component 1 (Fig. 3).

To determine which sequence orientation of the BBTV DNA component 1 ssDNA was present in virions, BBTV ssDNA was extracted from purified virions, electrophoresed through agarose and transferred to nylon membranes. These membranes were incubated with one of two 32P end-labelled oligonucleotides (primer A or B). It was found that primer A hybridized with virion ssDNA (Fig. 4a) whereas primer B did not (Fig. 4b). Primer A was complementary to the sequence presented in Fig. 3 confirming that this was the orientation present in virions.

The sequence of component 1 was analysed using the GCG program (Devereux et al., 1984). Three ORFs were found which could encode proteins of approximately 10K or greater (Fig. 5). ORF-V1 occurred in the virion orientation and had 858 nucleotides. This ORF contained a start codon at nucleotide 129 and terminated with a stop codon (TGA) at nucleotide 987. A poly(A) signal (AATAAA) was present from nucleotides 968 to 973. When translated, ORF-V1 potentially encoded a protein of 33-6K (Fig. 3). Upstream from the ORF-V1 start codon, there was one possible TATA box (TATAAA) from nucleotides 79 to 84. ORF-C1 (366 nucleotides) occurred in the complementary orientation from nucleotides 628 to 263 (Fig. 3). It potentially encoded a protein...
Fig. 5. Schematic representation of BBTV component 1 ORFs potentially coding for proteins of approximately 10K or greater: ORF-V1 occurs in the virion orientation whereas ORFs C1 and C2 occur in the complementary orientation. Positions of the stem-loop sequence and the potential ORF-V1 TATA box are indicated.

of 10K. No obvious poly(A) signal was associated with this ORF nor did there appear to be a 5’ TATA box. ORF-C2 (249 nucleotides) was also present in the complementary orientation from nucleotides 414 to 166. This ORF potentially encoded a protein of 9.3K, it did have a possible 5’ TATA box (from nucleotides 496 to 491) but there was no poly(A) signal associated with it.

The sequence was also analysed for possible stem–loop structures. A strong potential stem-loop occurred from nucleotides 28 to 58 containing a 10 bp stem and an 11 nucleotide loop (Fig. 3).

Discussion

We have demonstrated that the virions of BBTV contain circular ssDNA of approximately 1.1 kb using a PCR strategy that would amplify dsDNA from BBTV virion ssDNA only if this template were circular. The advantages of this strategy were (i) it was an efficient method for generating linear, full-length dsDNA using BBTV virion ssDNA as a template in a form suitable for cloning, (ii) only a small amount of sequence information was required from anywhere within the BBTV component and (iii) if the resultant amplified product was of the expected BBTV component size, the presence of this product established the circularity of that component. This method should be useful for generating full-length dsDNA copies of other potential BBTV components, other ssDNA viruses such as SCSV and CFDV, geminiviruses and PCV and related viruses and perhaps other small circular DNA molecules, the size being limited by the efficiency of the PCR amplification.

We have sequenced one component, component 1, of the BBTV DNA genome from both the original cDNA clone (pBT338) and three full-length PCR clones. There was a strong stem–loop structure predicted in the virion orientation of component 1; the loop sequence of 11 nucleotides contained a nine nucleotide sequence (TAATATTAC) which was almost identical to the invariant loop sequence present in nine geminiviruses (TAATATTAC) (Lazarowitz, 1987) and also CFDV (TAGTATTAC) (Rohde et al., 1990), with only one nucleotide difference in each case. Evidence from the study of geminiviruses indicates that this sequence is involved in DNA replication (Revington et al., 1989). The stem sequence of BBTV component 1 varied from that of CFDV and the geminiviruses.

Component 1 contained three ORFs that potentially encoded proteins of approximately 10K or greater. The largest ORF (ORF-V1) occurred in the virion sense and potentially encoded a replicase as it contained the dNTP-binding motif G(GE)GKT. The G(X)GKT motif has been shown to be associated with both RNA and DNA virus replicases (Gorbalenya et al., 1990). BBTV ORF-V1 was similar to the largest ORF (ORF1) in the component of CFDV that has been sequenced (Rohde et al., 1990). Both ORFs were in the virion orientation with start codons 3’ of the predicted stem–loop sequence; both ORFs had poly(A) signals starting 19 nucleotides 5’ of the stop codon and possible TATA boxes 5’ of the start codon (BBTV, TATAAA; CFDV, TATAAG); both ORFs potentially encoded proteins about 33K (BBTV ORF-V1, 33.6K; CFDV ORF1, 33.4K) and both these proteins had dNTP-binding motifs (BBTV, GGEK; CFDV, GGDGKS) starting at amino acid positions 183 and 184 respectively (Fig. 6). The two derived amino acid sequences were compared after alignment with the GCG PileUp program: there was 33% sequence similarity over the 286 amino acids of BBTV ORF-V1 with 47% sequence similarity over the 104 carboxy-terminal amino acids from the dNTP-binding motif. Conversely, there were no ORFs in the CFDV sequence that corresponded to BBTV ORF-C1 and ORF-C2 and no significant sequence similarity could be detected between these two BBTV ORFs and any CFDV ORF either at the nucleotide or amino acid level. Furthermore, a computer search failed to reveal any significant sequence similarity between these two BBTV ORFs and any published nucleotide or protein sequence. This would suggest that either BBTV component 1 and the CFDV component have different genome organizations or that BBTV component 1 and the CFDV component contain only one gene.

Only one recognized plant virus group has ssDNA as its genomic material, the geminiviruses. However, BBTV differs from the geminiviruses in a number of important
characteristics: BBTV has isometric virions (geminiviruses have geminate virions), BBTV is transmitted by aphids (geminiviruses are transmitted by leafhoppers or whiteflies), the unit size of the BBTV genome is about 1.1 kb (geminivirus virion DNA is about 2-7 kb) and BBTV has a coat protein of about 20K (geminivirus coat proteins are 26K to 34K). BBTV is more similar to CFDV (Harding 1987). Almost certainly, the BBTV isometric virions containing circular ssDNA of about 1 kb. SCSV is also transmitted by aphids, but CFDV is transmitted by Myndus taffini, a planthopper. Furthermore SCSV has a coat protein of about 19K; the size of CFDV coat protein has not been reported. The sequence of one component of CFDV has been determined (Rohde et al., 1990); it is not known at this time whether CFDV has a mono- or multi-component genome. It has been reported that the genome of SCSV consists of seven components of ssDNA (Chu et al., 1990). Almost certainly, the BBTV genome consists of more than one component: BBTV has a coat protein with an approximate M, of 20100 (Harding et al., 1991; Thomas & Dietzgen, 1991) and no ORF that could potentially encode a protein of this size was found in component 1. We are therefore attempting to clone and sequence further components of the BBTV genome and investigate whether mRNAs are transcribed in vivo from one or all of the ORFs of component 1.

From the evidence presented here, it would appear that BBTV belongs to an undescribed group which could include SCSV and CFDV. Three animal viruses, CAV, PBFDV and PCV, also have small isometric virions (17 to 22 nm) that contain circular ssDNA. These three viruses form the family Circoviridae; BBTV, CFDV and SCSV are also potentially members of this family. There are however some important differences between these animal and plant viruses. The virion ssDNA of the animal viruses is apparently monopartite and 1-7 to 2-3 kb in size (Todd et al., 1991); CAV and PCV have one virion-associated protein of 50K and 36K respectively and PBFDV has three virion-associated proteins of 15-9K, 23-7K and 26-3K (Ritchie et al., 1989). Furthermore, there was no significant sequence similarity between BBTV component 1 and CAV (Noteborn et al., 1991) either at the nucleotide or at the amino acid level.

We thank Mr Eric Gall and Mr Don Andersen of the Queensland Department of Primary Industries for BBTV-infected banana plants. This project was supported by the Australian Research Council and the Rural Industries Research and Development Corporation.

References


(Received 25 August 1992; Accepted 22 October 1992)