The genome organization of banana bunchy top virus: analysis of six ssDNA components

Thomas M. Burns,† Robert M. Harding and James L. Dale*

Centre for Molecular Biotechnology, School of Life Science, Queensland University of Technology, GPO Box 2434, Brisbane, Queensland 4001, Australia

We have cloned, sequenced and analysed an additional five circular ssDNA components of banana bunchy top virus (BBTV) which we have called components 2, 3, 4, 5 and 6. These components were present in all BBTV infections tested. Four of these components (components 3, 4, 5 and 6) had one large open reading frame (ORF) in the virion sense located 3' of a stem-loop structure. Each ORF had a potential TATA box and one or two potential polyadenylation signals associated with it and each polyadenylation signal had an associated GC-rich region containing the trinucleotide sequence TTG. A number of ORFs were identified in component 2 but none of these had appropriately located potential TATA boxes and polyadenylation signals associated with them. None of the ORF amino acid sequences nor the full DNA sequences of any of the components had significant sequence identity with any known protein or nucleic acid sequences. However, the ORF of component 4 encoded a 30 residue hydrophobic domain which may indicate that this ORF encoded a transmembrane protein. Further, the ORFs of components 3 and 5 potentially encoded proteins of about 20 kDa, the size of the BBTV coat protein. There were two regions of sequence identity between the five components described here and the previously described component 1. Each component contained a conserved stem-loop structure and a nonanucleotide potential TATA box which was 5' of the large virion-sense ORF in five of the components. The stem-loop structures were incorporated in a common region (CR-SL) of 69 nucleotides which was 62% identical between components. All six BBTV components also contained a major common region (CR-M) which was located 5' of the CR-SL in each component, in the non-coding region and was 76% identical over 92 nucleotides. Each CR-M contained a near-complete 16 nucleotide direct repeat and a GC-box which was similar to the rightward promoter element found in wheat dwarf geminivirus. From these results, BBTV appears to belong to an undescribed plant virus group which could also include subterranean clover stunt virus, coconut foliar decay virus, faba bean necrotic yellows virus and milk vetch dwarf virus.

Introduction

Banana bunchy top disease (BBTD) is the most important virus disease of bananas (Dale, 1987). The disease is widespread in Asia and the South Pacific and has limited distribution in Australia and Africa. It has not been reported from the Americas. The disease was originally assumed to be caused by a luteovirus as it was persistently aphid transmitted but not mechanically transmitted, induced yellows type symptoms and infected plants had damaged phloem. However, recently 18–20 nm isometric virus-like particles (VLPs) have been purified from infected plants and have been demonstrated to be associated with the disease (Harding et al., 1991; Thomas & Dietzgen, 1991; Wu & Su, 1990). Harding et al. (1991, 1993) have isolated circular, ssDNA of about 1 kb from these VLPs and cloned and sequenced one ssDNA component. This component, known as banana bunchy top virus (BBTV) DNA component 1, had one large open reading frame (ORF) in the virion sense and encoded a putative replicase. This component was transmitted with the disease via aphids. Karan et al. (1994) have demonstrated that BBTV DNA component 1 is present in all BBTD-infected plants tested from 10 different countries and provided evidence for two groups of isolates, the Asian group including isolates from the Philippines, Taiwan and Vietnam and the South Pacific

* Author for correspondence. Fax +61 7 864 1534; e-mail dalej@qut.edu.au

† Present address: Virus Research Department, John Innes Institute, Colney Lane, Norwich NR4 7UH, UK.

The sequence data reported in this paper have been deposited in GenBank and assigned the accession numbers L41574 to L41578.
group including isolates from Australia, Burundi, Egypt, Fiji, India, Tonga and Western Samoa. Yeh et al. (1994) reported the sequence of a second BBTV ssDNA component; however, the sequence of this component was derived from two overlapping clones and a large ORF could not be resolved.

More recently, we have demonstrated that the genome of BBTV comprises at least five ssDNA components based on the restriction digests of putative full-length clones synthesized from purified virions (Burns et al., 1994). Thus, BBTV is a multicomponent virus and appears to be most similar to subterranean clover stunt virus (SCSV) which has small isometric virions and a genome comprising of at least seven circular ssDNA components each of about 1 kb (Chu et al., 1990, 1993). Both of these viruses are persistently aphid transmitted and have coat proteins of about 20 kDa and probably form a new group of plant viruses. This group could probably also include the aphid transmitted faba bean necrotic yellows virus (FBNYY; Katul et al., 1993) and milk vetch dwarf virus (MVDV; Sano et al., 1993), and possibly the plant hopper transmitted coconut foliar decay virus (CDFV; Rohde et al., 1990).

In this paper, we report the sequence and analysis of five new ssDNA components of BBTV and discuss the relationship of BBTV with other ssDNA viruses.

Methods

Synthesis and cloning of cDNA. Bananas with characteristic symptoms of BBTD were collected from the Nambour region of southeast Queensland. BBTV particles were purified as described by Wu & Su (1990) and Thomas & Dietzgen (1991). Nucleic acid was extracted from virions as described by Francki & Randles (1973); dsDNA was synthesized as described by Gubler & Hoffman (1983) using random hexamers (Bresatec) to prime first strand synthesis. The dsDNA was treated with mung bean nuclease (Promega) and ligated using 0.4 M-NaOH (Sambrook et al., 1989). Inserts were excised by digestion with EcoRI-HindIII, electrophoresed in agarose gels and capillary blotted onto nylon membrane. The plasmid was then used to transform Escherichia coli strain JM109 (Hanahan, 1983) and potential recombinant clones were identified by screening on X-Gal substrate (Vieira &Messing, 1982).

Plasmids were isolated using the alkaline lysis method (Sambrook et al., 1989). Inserts were excised by digestion with EcoRI-HindIII, electrophoresed in agarose gels and capillary blotted onto Hybond-N+ (Amersham) using 0.4 M-NaOH (Sambrook et al., 1989). Inserts for use as DNA probes were purified from agarose gels using a Gene-Clean kit (Bresatec). DNA probes were labelled using a Ready-To-Go labelling kit (Pharmacia) as recommended by the manufacturer. Prehybridizations and hybridizations were done as described by Burns et al. (1994).

Sequencing and sequence analysis. Mini-preparations of respective BBTV clones were prepared by alkaline lysis followed by polyethylene glycol precipitation (Hattori & Sakaki, 1986). Sequencing was done using [35S]dATP and a Sequenase kit (USB) as recommended by the manufacturer. Reaction products were electrophoresed in 8% (w/v) polyacrylamide gels containing 7 M-urea. Gels were fixed, dried and exposed to X-ray film. Primers used for sequencing were either universal sequencing primers or 17–30 nucleotide primers complementary to appropriate regions of the cloned viral DNA synthesized using an Applied Biosystems (ABI) PCR Mate and processed as recommended by the manufacturer.

PCR products for sequencing were purified from agarose gels using a Gene-Clean kit (Bresatec). DNA was sequenced using a Sequenase kit (USB) essentially as described by the manufacturer. Denaturation of template DNA (500 ng) was done by boiling following the addition of DMSO and 3 pmol of sequencing primer.

Nucleotide sequences were analysed using the GCG analysis package version 8, available through the ANGIS computing facility at the University of Sydney, Australia. Nucleotide and amino acid sequences were aligned using the Clustal V software package (Higgins et al., 1991). Four DNA databases (GenBank, GenBank Weekly Updates, EMBL and EMBL weekly updates) and five protein databases (SwissProt, SwissProt Weekly Updates, PIR, GenPeptide Proteins and GenPeptides Weekly Updates) were searched for sequence homologies with BBTV nucleotide and deduced amino acid sequences using two database search analysis programs, FASTA (Pearson & Lipman, 1988) and BLAST (Altschul et al., 1990).

PCR: Analysis and cloning. Using the respective nucleotide sequences of clones (i) pBTRP-11, 20, 80 and 88 and (ii) pBTRP-P1 and P2 and the nucleotide sequences of BBTV components 1 (Harding et al., 1993) and 2, three sets of immediately adjacent outwardly extending primers (i) primer A, 5’ GATCGCAAGGCAGCTA3’; primer B, 5’ CTCCCCGGAAGCTTGG3’; (ii) primer C, 5’ TATAGTACGGACGACTA3’; primer D, 5’ TAACTTCCAGTCTCTT3’; (iii) primer E, 5’ CGGGGTATATGAAATTGGT3’; primer F, 5’ TGGAGGTTTCTGG3’; and primer G, 5’ TTTGTGCTACGCTGT3’ were synthesized and used in a PCR with BBTV DNA as template as described by Burns et al. (1994). The amplified products were cloned using the TA cloning kit (Invitrogen) into the plasmid vectors pCR1 or pCR2000 as recommended by the manufacturer or into T-tailed pUC19 and Bluescript (Marchuk et al., 1990). Recombinant clones were selected using X-Gal substrate on Luria–Bertani (LB) agar containing the appropriate antibiotic and plasmids isolated using the alkaline lysis method (Sambrook et al., 1989). Clones with apparent full-length inserts (approximately 1 kb) were selected for sequencing.

Polarity of virion ssDNA. BBTV ssDNA was extracted, electrophoresed in agarose and capillary blotted onto duplicate nylon membranes (Harding et al., 1993). For component 2, a DNA 3’ end-labelling kit (Boehringer Mannheim) was used to prepare 32P-labelled strand-specific oligonucleotide hybridization probes [primer BT2RF30 (G), 5’ GTGTTTCTTTTTTATTCCGGC3’; primer BT2RF30 (H), 5’ CGGAAAAATAGAATATAGTGGGAGTGAGT3’]. Membranes were prehybridized and hybridized for 12 and 20 h respectively in Rapid-hyb (Amersham) at 60 °C. Filters were washed once with 1% SDS, 2 x SSC (0.3 M-NaCl, 0.03 M-sodium citrate, pH 7.0) at room temperature followed by washes with 2 x SSC at 65 °C. Dried membranes were exposed to X-ray film at −80 °C using intensifying screens.

For components 3, 4, 5 and 6 strand-specific RNA probes were used. Full-length RNA transcripts of full-length BBTV clones of each of the four components were synthesized using a riboprobe in vitro transcription kit (Promega) as recommended by the manufacturer.

Results

Cloning and sequencing of five genomic components

Five new genomic components of BBTV were cloned and sequenced from two libraries, (i) a random primed library and (ii) a PCR library.
(i) Random primed library

A random primed library was generated from BBTV ssDNA extracted from purified virions. The resultant dsDNA was treated with mung bean nuclease, blunt-end ligated into SmaI cut pUC18 and cloned into E. coli JM109. This library was screened with 32P-labelled DNA from BBTV virions, healthy bananas and the insert from pBT338 which was a partial clone of BBTV DNA component 1 (Harding et al., 1991, 1993).

**BBTV DNA component 2.** Four clones, pBTRP-11, 20, 80 and 88, hybridized with BBTV virion DNA and with each other but not with healthy banana DNA or pBT338. The inserts from these clones were sequenced: pBTRP-20 and 88, each with inserts of 220 bp, had identical sequences; pBTRP-80 had an insert of 188 bp and had 148 bp sequence identical to pBTRP-20 and 88 and a further 40 bp sequence at one end that was unique; the sequence of the 115 bp insert of pBTRP-11 was identical to the equivalent region of pBTRP-20 and 88. Two immediately adjacent, outward extending primers, primers A and B, were designed from the overlapping sequence of the four clones such that these primers would prime the amplification of full-length dsDNA copies of a circular ssDNA molecule (Harding et al., 1993). BBTV virion ssDNA was amplified by PCR using these primers and Pfu DNA polymerase and the product cloned into pCR2000. Four of the resultant clones were sequenced in both directions using universal forward and reverse primers and sequence specific primers. Three of these clones contained 1060 bp inserts and one clone contained a 1059 bp insert. The four clones had identical sequences except for nine single nucleotide changes including one deletion. Further, the sequences of the original four cDNA clones were found within the four PCR clones. The consensus sequence of this component, termed BBTV DNA component 2, was compiled (Fig. 1a) and compared with the sequence of BBTV DNA component 1 (Harding et al., 1993); the two sequences were essentially different apart from two significant regions of identity.

**BBTV DNA component 6.** A further two clones from the same random primed library, pBTRP-P1 and P2, also hybridized with labelled BBTV virion DNA but not with DNA from healthy bananas or pBT338. However, the inserts of these clones, both of approximately 1 kb, were digested with EcoRV whereas neither components 1 nor 2 had EcoRV sites. The two clones were partially sequenced using universal forward and reverse primers. The sequences of both clones were identical but clearly different to those of components 1 and 2. Again, two immediately adjacent, outwardly extending primers, primers C and D, were designed from the sequence and synthesized. BBTV virion ssDNA was used as a template with these two primers in a PCR reaction and the resultant product cloned into a T-tailed Bluescript vector. One apparent full-length clone, pBT-P2A1, was selected and sequenced in both directions from subclones generated by exonuclease III digestion and universal forward and reverse, and sequence-specific primers. The final component 6 sequence of 1089 bp was then compiled (Fig. 1e).

(ii) PCR library

When the sequences of components 1 and 2 were compared, two regions of identity were identified. The first region, later defined as the stem–loop common region (CR-SL) included the potential stem-loop sequence previously identified in component 1 (Harding et al., 1993); the second region, which was contained within the region later defined as the major common region (CR-M), was a sequence of approximately 66 nucleotides 5' to the stem–loop sequence. It was hypothesized that all BBTV genomic components should contain a CR-M and therefore two immediately adjacent, outwardly extending degenerate primers, primers E and F, were designed from this region, synthesized and extended by PCR using BBTV virion ssDNA as a template (Burns et al., 1994). Seven products, each of approximately 1 kb, were resolved by polyacrylamide gel electrophoresis. The products were cloned into pCRII. The resultant clones were divided into three groups, groups B, C and D, on the basis that they hybridized with BBTV virion DNA but not DNA from healthy bananas and that each group had restriction patterns different to the other two groups and to components 1, 2 and 6 (Burns et al., 1994). One group, group A, had a restriction pattern indistinguishable from that of component 2 and it was later confirmed by sequencing that group A clones represented clones of component 2.

Each group of clones was assumed to represent a new and unique BBTV DNA component. For each group of clones, three clones (component 3) or four clones (components 4 and 5) were partially sequenced using universal forward and reverse primers. In each instance, all the clones within a group had identical sequences where these sequences overlapped except for one or two single nucleotide changes. Further, the sequences from each group were different to each other group and to the sequences of components 1, 2 and 6. One clone from each group or component was selected and fully sequenced in both directions. Importantly, each of these groups of clones were generated using degenerate primers covering a sequence of 34 nucleotides derived from the conserved CR-M of components 1 and 2. The CR-M from
components 1 and 2 was not fully conserved and thus it was expected that the hypothesized CR-M sequence would vary between other components. Therefore, converging primers unique to each component were designed and used to amplify a sequence including CR-M for each component from BBTV virion ssDNA. The resultant PCR product was sequenced directly using the two component-specific converging primers.

**BBTV DNA component 3.** Component 3 (Group C clone pBTP-64) was sequenced in both directions from the original clone and from subclones generated by exonuclease III digestion or restriction fragments using universal forward and reverse primers and three sequence-specific primers. Two additional converging primers were designed from this sequence to amplify a 380 bp product including the CR-M. The sequence of this product was identical to that of pBTP-64 except for five single nucleotide changes, four of which were in the sequence covered by the original degenerate primers and
one outside this sequence at nucleotide 947. The final component 3 sequence of 1075 bp was then compiled (Fig. 1b).

**BBTV DNA component 4.** Component 4 (Group D clone pBTP-62) was sequenced in both directions from the original clone and from subclones generated by exonuclease III digestion using universal forward and reverse primers and three sequence-specific primers. Two additional converging primers were designed from this sequence to amplify a 350 bp product including the CR-M. The sequence of this product was identical to that of pBTP-62 except for two single nucleotide changes in the sequence covered by the original degenerate primers. The final component 4 sequence of 1043 bp was then compiled (Fig. 1c).

**BBTV DNA component 5.** Component 5 (Group B clone pBTP-129) was sequenced in both directions from the original clone and from subclones generated by exonuclease III digestion using universal forward and reverse primers and three sequence-specific primers. Two additional converging primers were designed from this sequence to amplify a 290 bp product including the CR-M. The sequence of this product was identical to that of pBTP-129 except for four single nucleotide changes in the sequence covered by the original degenerate primers. The final component 5 sequence of 1018 bp was then compiled (Fig. 1d).

### Orientation of genomic components and association with banana bunched top disease

We have previously shown that the BBTV genome is encapsidated as ssDNA (Harding et al., 1991, 1993). To determine the orientation of each component in virions, strand-specific DNA or RNA probes specific for each component were synthesized and hybridized with BBTV virion DNA. Component 2 specific probes were two 3' end-labelled 30-mer oligonucleotides whereas probes specific for components 3, 4, 5 and 6 were SP6, T3 or T7 promoted 32P-labelled RNA transcripts. For each component, the probes whose sequences were complementary to the component sequences presented in Fig. 1 hybridized strongly to BBTV virion DNA whereas the probes whose sequences were the same as the Fig. 1 sequences did not hybridize (Fig. 2). This indicated that each component was encapsidated as ssDNA and only in one orientation, that presented in Fig. 1.

Further, the strand- and component-specific probes that hybridized with BBTV virion ssDNA were used as probes to demonstrate that each component was associated with BBTD. Plant DNA extracts from three (for component 2) or four (for components 3 to 6) different BBTV isolates and DNA from four healthy bananas were Southern blotted and hybridized with each probe. Each component-specific probe hybridized with a low molecular mass DNA of the expected size in all the extracts from BBTV-infected bananas but did not hybridize to DNA from healthy bananas.
hybridize with the extracts from healthy bananas (results not shown). This indicated that each component was clearly associated with the disease and the virus.

Analysis of the BBTV genomic components

The sequences of the five BBTV genomic components presented here and the sequence of component 1 (Harding et al., 1993) were aligned and compared. Each of the six sequences were different except for two significant regions which had varying degrees of identity between all six components.

Stem–loop common region. We have previously identified a potential stem–loop structure in BBTV component 1 (Harding et al., 1993) which had a loop sequence almost identical to the invariant loop sequence of geminiviruses (Lazarowitz, 1992). An equivalent stem–loop structure was also found in BBTV components 2 to 6 (Fig. 3). Each component had an 11 nucleotide loop sequence of which nine consecutive nucleotides were conserved between all components. Each component also had a 10 bp stem sequence of which 14 nucleotides were fully conserved. However, when all six components were compared, the region of identity extended up to 25 nucleotides 5' of the stem–loop structure and up to 13 nucleotides 3' of the stem–loop structure. The 5' 25 nucleotides were fully conserved between components 1, 3, 4 and 5. There were apparently two deletions in both components 2 and 6. In component 2, eight nucleotides were fully conserved with components 1, 3, 4 and 5 whereas in component 6, 16 nucleotides were conserved with these other components. The 13 nucleotides 3' of the stem–loop were fully conserved between all six components except for an apparent single nucleotide deletion in component 2. The sequence of up to 69 nucleotides including the stem–loop sequence was termed the stem–loop common region or CR-SL.

Major common region. The second common region was located at various distances 5' of the CR-SL and was called the major common region or CR-M. This region

Fig. 2. Determination of the virion-sense orientation of BBTV DNA components 2 to 6. Each blot was separately probed with either 32P-labelled oligonucleotides (component 2) or full-length RNA transcripts (components 3 to 6) specific for the virion- or complementary-sense strands of each respective component. (a) Blots hybridized with probes complementary to the component sequences presented in Fig. 1; (b) blots hybridized with probes of the same sequences presented in Fig. 1. Lane 1, full-length clone of each respective component; lane 2, healthy banana nucleic acid; lane 3, DNA extracted from purified BBTV virions.

Fig. 3. Aligned stem-loop common regions (CR-SL) of BBTV DNA components 1 to 6. The stem–loop structure in each component is underlined and the loop sequence is in italics. Asterisks indicate nucleotides that are conserved between all components. Dots have been included in some sequences to maximize sequence alignment.
varied in size from 65 nucleotides in component 1 to 92 nucleotides in component 5 (Fig. 4). Component 1 apparently had the first 26 nucleotides of the CR-M deleted as well as a further single nucleotide deletion. Components 2, 3 and 4 had two single nucleotide deletions and component 6 had one single nucleotide deletion. Forty-five nucleotides were conserved between all components and 23 of the first 26 nucleotides, deleted as well as a further single nucleotide deletion. Component 1, were conserved between components 2 to 6. Also in components 2 to 6, there was an almost complete 16 nucleotide direct repeat (ATACAAc/gAc/gCTATG) from nucleotides 4 to 20 and 21 to 36. Further, a 15 nucleotide GC-rich sequence (average of 86% G + C) was located from nucleotides 78 to 92 and was 93% conserved between all components.

The sequence between the last nucleotide of the CR-M and the first nucleotide of the CR-SL varied in length from 22 nucleotides in component 1 to 233 nucleotides in component 2 (Figs 1 and 5). Interestingly, this sequence of 175 nucleotides in components 3 and 4 was 97% conserved between these two components.

Potential TATA boxes. A potential TATA box was identified in BBTV component 1 and was located 20 nucleotides 3' of the last nucleotide of the stem–loop and 43 nucleotides 5' of the start codon of the putative replicase gene (Harding et al., 1993). Similar potential TATA boxes were also identified in components 2 to 6. In each of these components, the potential TATA box was a nine nucleotide sequence, CTATa/ta/tAt/aA, and was located downstream from the stem–loop sequence (Fig. 1). However, the sequence between the last 3' nucleotide of the stem–loop sequence and the potential TATA box was considerably longer in components 2 to 6 than in component 1 and varied from 157 nucleotides in component 5 to 227 nucleotides in component 4 (Figs 1 and 5).

Analysis of components for open reading frames

One large ORF coding for a putative replicase was identified in the virion sense of BBTV component 1 which had a potential TATA box 43 nucleotides 5' of the ATG start codon and a polyadenylation signal 13 nucleotides 5' of the stop codon (Harding et al., 1993). Components 2 to 6 were therefore analysed for ORFs in both the virion and complementary sense that could code for proteins of more than 25 amino acids. Numerous such ORFs were identified in both orientations in all five components. However, only four ORFs were identified that had associated with them a potential 5' TATA box and an appropriately located polyadenylation signal. Components 3 to 6 each had one such ORF in the virion sense. These ORFs were (i) 525 nucleotides potentially coding for a 175 amino acid protein of 20.11 kDa in component 3, (ii) 351 nucleotides potentially coding for a 117 amino acid protein of 13.74 kDa in component 4, (iii) 483 nucleotides potentially coding for a 161 amino acid protein of 18.97 kDa in component 5 and (iv) 462 nucleotides potentially coding for a 154 amino acid protein of 17.4 kDa in component 6 (Figs 1 and 5).

Nine ORFs were identified in component 2, four in the virion sense and five in the complementary sense. However, none of these ORFs had appropriately located nonanucleotide potential TATA boxes and polyadenylation signals and therefore were unlikely to be transcribed.

Analysis of potential polyadenylation signals

Six potential polyadenylation signals were identified associated with the 3' end of the major ORFs of components 3 to 6. A GT-rich region of 10 to 17 nucleotides was located between 0 and 23 nucleotides 3' of each of these polyadenylation signals and each GT-rich region contained the trinucleotide sequence TTG (Fig. 1). Only one potential polyadenylation signal identified in component 2 had a corresponding GT-rich region with the trinucleotide sequence TTG and this was located 233 nucleotides 3' of the nonanucleotide potential TATA box in the virion sense.

Sequence comparisons and analysis

Four DNA databases were searched using the complete nucleotide sequences of components 2 to 6 and five protein databases were searched using the putative amino acid sequences derived from the major ORFs of
Fig. 5. Diagrammatic representation of the proposed genome organization of BBTV. (a) General organization of all components; (b) linear representation of each component.
components 3 to 6 as well as the smaller ORFs. No significant sequence identity was found either at the nucleotide or amino acid level. Further, no motifs or other signals were identified that would suggest possible functions for any of the four putative proteins other than 30 hydrophilic residues toward the N terminus of the major ORF translation product of component 4 suggesting the presence of a transmembrane domain (Boulton et al., 1993). The ORFs of components 3 and 5 encoded putative proteins with molecular masses close to the 20.1 kDa molecular mass of the BBTV coat protein. However, there is no other evidence to suggest either of these ORFs encoded the coat protein.

Discussion

We have cloned and sequenced a further five ssDNA components of the BBTV genome. Including BBTV component 1 (Harding et al., 1993), all six components shared two common regions, the CR-SL and the CR-M, in the putative intergenic or untranslated region and five of the six components had one large ORF in the virion sense with associated potential TATA boxes and polyadenylation signals (Fig. 5). The CR-SL incorporated the conserved stem–loop structure. The loop sequence of 11 nucleotides was conserved in all BBTV components with the exception of two nucleotides and was similar to that present in nine geminiviruses (Lazarowitz, 1992), CFDV (Rohde et al., 1990) and a further BBTV component (Yeh et al., 1993). A model for implicating the loop sequence in rolling circle replication has been described for geminiviruses (Saunders et al., 1993). It is possible that the loop sequence in BBTV has a similar function. The stem–loop sequences were also highly conserved in all BBTV components and contained the pentanucleotide sequence TACCC which has been shown to be the site for initiation of viral strand DNA synthesis in wheat dwarf geminivirus (Heyraud et al., 1993).

The major common region (CR-M) was identified in all components and was located 3' of the major ORF (except for component 2 where no major ORF was identified) and 5' of the CR-SL (Fig. 5). Hexanucleotide repeats were identified within the CR-M in all components except that of component 1. However, no function could be directly attributed to these repeats but they may be associated with, or part of promoter sequences. The CR-M also contained a 15 nucleotide GC-rich sequence located at the 3' end and had the potential to form a small stem–loop structure. This GC-rich sequence also contained two direct GC-repeats which resembled the Spl binding sites found in promoters of genes in animal cells and viruses (Fenoll et al., 1990). A similar promoter in the monocot-infecting maize streak geminivirus has been shown to be required for maximal rightward transcription and also appeared to bind maize nuclear factors in a non-cooperative manner (Fenoll et al., 1990).

Karan et al. (1994) reported that the component 1 CR-M sequence was highly conserved within the South Pacific group of BBTV isolates (96.5% identity) and within the Asian group of isolates (98.0% identity) but was highly variable between the two groups of isolates (68.0% identity). There was 76% identity between the CR-M sequences of the six different components of an Australian isolate reported here. Therefore, it will be important to determine the level of identity between the CR-M sequences of individual components from the different groups of isolates to see whether these sequences are highly conserved within groups of isolates but variable between groups and different components and further whether this has any biological significance.

The nucleotide length and sequence between the CR-M and CR-SL was dissimilar in four of the six components. However, in components 3 and 4, this 175 nucleotide region was 97% identical and the 334 nucleotides from the 5' end of the CR-M to the 3' end of the CR-SL were 98% identical. A similar large common region of 300 nucleotides has been found in geminiviruses and is identical between the A and B components of individual bipartite geminiviruses (Lazarowitz, 1992). In geminiviruses, this region included the stem–loop region. A region of identity which included the stem–loop region was also found in five of the seven components of SCSV (Surin et al., 1993); this region is similar to that in the geminiviruses but different to four of the six BBTV components.

Components 3, 4, 5 and 6 all had one large ORF in the virion sense, 3' of the CR-SL. Each of these ORFs had potential conserved TATA boxes and polyadenylation signals associated with them (Fig. 5). The potential TATA boxes are highly conserved with the nona-nucleotide sequence CTATA/ta/taA/ta which was essentially similar to that described by Bucher et al. (1990). The distance between the potential TATA box and the translation initiation codon varied in each component from 13 nucleotides in component 3 to 102 nucleotides in component 1. An ATGG translation initiation codon was identified in the five components encoding large ORFs. However, two possible translation initiation codons were identified in component 3, the first at nucleotide 213 (ATGT) and the second at nucleotide 227 (ATGG); the second initiation codon was in frame with the first. This would suggest that the second initiation codon is the correct codon; this could be verified by 5' RACE or N-terminal sequencing of the ORF translation product. GT-rich regions were identified 0 to 24 nucleotides 3' of each of the polyadenylation signals in components 1, 3, 4, 5 and 6. Each of these GT-rich
regions contained the nucleotide sequence TTG. Both the polyadenylation signals in components 3 and 6 had these sequences. The combination of a consensus polyadenylation signal (Aa/TAAa/t) and a 3' proximal GT-rich region containing the trinucleotide sequence TTG was only associated with the single major virion-sense ORF in components 1, 3, 4, 5 and 6 and was not identified elsewhere in these sequences, suggesting that each of these components encoded a single gene. Similar sequences have been associated with many polyadenylation signals and may be required for efficient termination (Gil & Proudfoot, 1984; Conway & Wickens, 1985).

None of the major ORFs of components 3, 4, 5 or 6 had significant sequence identity either at the DNA or protein level with any other available sequences and no functions could be assigned to the putative proteins. However, the ORFs of components 3 and 5 encoded proteins of a size similar to that of the BBTV coat protein, 20.1 kDa (Harding et al., 1991). Further, a sequence of 30 hydrophobic amino acids was identified near the N-terminal end of the putative protein of the component 4 ORF which is characteristic of in- or transmembrane domains. Similar hydrophobic domains have been identified in the proposed movement proteins of cereal-infecting geminiviruses but have not been identified in any other virus movement proteins including those of dicot-infecting geminiviruses (Boulton et al., 1993). No other ORFs were identified in components 3 to 6 in either virion or complementary sense which potentially encoded proteins greater than 10 kDa and had appropriately located potential TATA boxes and polyadenylation signals.

In component 2, sequence conforming to the consensus polyadenylation signal was identified 17 nucleotides 5' of a GT-rich region containing a TTG sequence. However, this potential polyadenylation signal was not associated with any major ORF in component 2. Further, no ORF was identified 3' of the nonanucleotide potential TATA box and 5' of the CR-SL as in the other five components. No other ORFs in either the virion or complementary sense of component 2 had appropriate potential TATA boxes or polyadenylation signals associated with them. Sequence analysis of component 2 derived from an Egyptian isolate and from BBTV-specific dsDNA purified from infected Australian bananas failed to identify such an ORF (A. Sadik, R. Wanitchakorn, R. Harding & J. Dale, unpublished results). Thus, while component 2 had a potential TATA box and polyadenylation signal location at the expected positions as extrapolated from the other five components, it is possible that this component does not encode a protein but may have a regulatory function.

We have demonstrated that each component was encapsidated as the single-stranded coding sense DNA and was present in all infected plants tested indicating that each of the five components is an integral part of the BBTV genome. We have named these as components 2 to 6. Yeh et al. (1994) reported a second component of BBTV and named this component 2. However, we now have evidence to suggest that this component is not an integral part of the BBTV genome (M. Karan, R. Harding and J. Dale, unpublished) and therefore should be re-named accordingly. Wu et al. (1994) have also isolated two BBTV-related DNA components, named DNA-1 and DNA-2. DNA-1 has high sequence similarity with the component reported by Yeh et al. (1994) whereas DNA-2 appears to encode a putative replicase or replicase associated protein but does not have significant sequence similarity with BBTV component 1 (Harding et al., 1993). It will be necessary to determine whether BBTV has two replicase associated proteins and whether both these components are necessary for infection.

Thus, BBTV has a genome of at least six encapsidated components of ssDNA about 1.0–1.1 kb and five of the six known components have a single large ORF in the virion sense. Such a genome organization is clearly different to any of the described plant virus groups and families. However, SCSV (Chu et al., 1990; Surin et al., 1993) has been reported to have a similar genome organization. This virus has seven known ssDNA components about 1 kb, two of which code for putative replicases. The sequences of these seven components have not been reported. We have identified only one putative replicase as an integral part of the BBTV genome. BBTV and SCSV share a number of other characteristics in that they have 18–20 nm isometric virions, a coat protein of 19–20 kDa, both appear to be phloem associated and are persistently aphid transmitted. Almost certainly, these two viruses will form the basis of a new plant virus group to which coconut foliar decay (Rohde et al., 1990), faba bean necrotic yellows (Katul et al., 1993) and milk vetch dwarf (Sano et al., 1993) may also belong.

It has been suggested that SCSV, BBTV and CFDV may be related to and included in a new virus family, the Circoviridae. The three animal viruses included in this family, porcine circovirus (PCV; Mankertz et al., 1993), chicken anaemia virus (CAV; Noteborn et al., 1993; Phenix et al., 1993) and psittacine beak and feather disease virus (PBFDV; Richie et al., 1989) have isometric virions (17 nm, 19–25 nm and 23–25 nm respectively) and a single component circular ssDNA genome (approximately 1.8 kb, 2.3 kb and 2.3 kb respectively). Six ORFs have been identified in the genome of PCV (Mankertz et al., 1993) and three protein products have been identified for CAV and PBFDV all of which are
translated from a single polyadenylated RNA transcript. PCV and CAV have coat proteins of 30 kDa and 50 kDa respectively. It would therefore appear that the only characteristics BBTV and SCSV share with the Circoviridae are that all have small isometric virions and a ssDNA genome. The genome organizations of these plant and animals viruses differ markedly and we feel that the inclusion of BBTV and SCSV in the Circoviridae is unjustified. It would be more appropriate to consider a new plant virus group that would include at least BBTV and SCSV.

We thank Peter Beetham and Greg Hafner for assistance with the preparation of figures and Eric Gall (Queensland Dept Of Primary Industries) for the supply of BBTV-infected material. This work was supported by a grant from the Australian Research Council. One of us (T.M.B) was supported by a Rural Industries Research and Development Corporation scholarship.

References


*(Received 8 November 1994; Accepted 14 February 1995)*