Processing of complementary sense RNAs of *Digitaria* streak virus in its host and in transgenic tobacco

Philip M. Mullineaux*, François Guerineau¹ and Gian-Paolo Accotto²

John Innes Institute, John Innes Centre for Plant Science Research, Colney Lane, Norwich NR4 7UH, ¹Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK and ²Istituto di Fitovirologia del CNR, Strada delle Cacce 73, 10135 Torino, Italy

Received October 1, 1990; Revised and Accepted November 19, 1990

ABSTRACT

We have used a polymerase chain reaction (PCR) procedure to analyse low abundance complementary sense RNAs of *Digitaria* streak virus (DSV) from infected leaves of *Digitaria setigera*. This study has confirmed that both spliced and unspliced RNAs are synthesised by the same transcription unit. The position of the intron has been proven from sequencing cDNAs corresponding to the spliced RNA. Although the majority of cDNAs have 3' ends at coordinate 1063, downstream from a consensus polyadenylation sequence, a minor population of RNAs with heterogeneous 3' ends has also been identified. Two major RNA species with alternative splice sites or 3' ends, previously identified by nuclease S1 protection assays, could not be detected, but a cDNA species was observed with an apparent 90bp insertion at the 5' end of the intron.

In transgenic tobacco containing integrated dimers of DSV DNA, the major unspliced RNA could readily be detected, but no spliced RNA was present. This may be a reason why DSV DNA did not replicate in tobacco. In addition, neither the minor population of heterogeneous RNAs nor the cDNA species with the insertion could be detected. The failure of the intron to be spliced in tobacco and its low activity in *Digitaria* is discussed in relation to recent studies on RNA splicing in plants and has led us to the conclusion that the geminivirus introns may be intrinsically inefficient.

INTRODUCTION

Digitaria streak virus (DSV) is a member of a sub-group of geminiviruses which infect members of the Gramineae (1) and may be good models to study aspects of gene expression in plants (2). Geminiviruses are characterised by the possession of genomes of circular single-stranded (ss) DNA encapsidated in twinned quasi-isometric particles (1). Geminivirus DNA has been isolated from infected plants as a supercoiled double-stranded (ds) form (reviewed in 3) as well as minichromosomes (4), which by

analogy with other DNA viral systems such as the papovaviruses (5) may represent transcriptionally active forms of the virus.

Nucleotide sequence data available for several members of the different sub-groups of geminiviruses (summarised in 1), including DSV (6,7) indicate that transcription of their genomes is bidirectional. This has been confirmed by mapping the transcripts of several geminivirus genomes, including African cassava mosaic virus (ACMV; 8), tomato golden mosaic virus (TGMV; 9), maize streak virus (MSV; 10) and DSV (2). These studies have also revealed substantial differences in the organisation of transcription of the genomes of those viruses which infect the Gramineae and those which infect dicotyledonous hosts. The most notable features of the RNAs which are encoded by the Gramineae infecting sub-group are;

1. Two 3' co-terminal transcripts, with differing 5' ends, which code for the virion sense open reading frames (V1 and V2 ORFs; see Fig 5 and 10; 2),
2. Rare spliced complementary sense RNAs, produced by excision of an intron located at the junction between the C1 and C2 ORFs which leads to the synthesis of a 41kDa C1:C2 polypeptide (see Fig. 1 and 5; 2 and 11).

Replication studies in protoplasts of *Triticum monoccocum* using a wheat dwarf virus (WDV) 'intronless' mutant, have shown that the C1:C2 fusion product is the only virus encoded product required for ds DNA replication (11). In addition, more abundant unspliced RNAs are made by the same transcription unit and direct the synthesis of the C1 product (2).

In addition to the major RNA species of DSV, detailed studies using the S1 nuclease protection assay (12) revealed three apparently minor, complementary sense RNAs (2), which had either a different 5' end (RNA 1-), or a different 3' end (RNA 3-) or an alternative splicing site (RNA 5-). Due to the very low abundance of the DSV complementary sense RNAs, the complexity of their 3' ends and the presence of two apparent internal S1 nuclease sites (only one of which has been correlated with the presence of consensus donor and acceptor sites) have made it necessary to analyse them by an alternative procedure which enables complementary sense cDNAs to be cloned and sequenced.

* To whom correspondence should be addressed

In this paper we report the use of the 3'RACE (rapid amplification of cDNA 3' ends) polymerase chain reaction (PCR) technique (13) to analyse complementary sense, DSV specific RNAs from infected *Digitaria* tissue and from transgenic tobacco plants containing an integrated tandem dimer of DSV DNA. Using this technique, we have characterised cDNAs corresponding to the spliced and unspliced RNAs, fine mapped their 3' ends, confirmed that they are polyadenylated and that only one intron was present. We have shown that in the transgenic tobacco plants, the major complementary sense unspliced transcript was the same as its counterpart from infected *Digitaria* plants. However, no spliced RNA could be detected.

MATERIALS AND METHODS

Materials

All restriction endonucleases and DNA modifying enzymes were purchased from Gibco-BRL with the exception of Amplitaq (Taq DNA polymerase) which was bought from Perkin-Elmer Cetus. ³²P-labelled nucleotides were purchased from New England Nuclear.

Plants and virus

Digitaria setigera L. was obtained from Vanuatu and grown under the same conditions as described for *Zea mays* L. (14). DSV was inoculated into *Digitaria* plants as pDS2 (7) using the 'agroinfection' technique (15). Tobacco (*Nicotiana tabacum* L. cv. Samsun) was grown as described previously (16).

RNA preparation, cDNA synthesis and PCR amplification

The preparation of total and poly(A)+ RNA from infected or healthy *Digitaria* leaf tissue has been described previously (2). The same procedure was used to prepare total RNA from young leaves of transformed and control tobacco plants.

First strand cDNA synthesis and PCR amplifications were carried out essentially as described by Frohman et al (13) using either 1 µg of poly(A)+ RNA or 20 µg of total RNA. The cDNA was diluted to a final volume of 1 ml (poly(A)+ RNA) or 200 µl (total RNA) using sterile distilled water. Control cDNA reactions were carried out using either 1 µg poly(A)+ RNA or 20 µg total RNA pretreated with 1 µg/ml DNAase free RNAase A for 60 minutes at 37°C. The primer used was the dT₁₇-adaptor (13); 5'-GACTCGAGTCGACATCGATTTTTTTTTTTTTTTT3').

The PCR amplification reactions were carried out in a final volume of 100 µl overlaid by 50 µl of mineral oil using a Perkin-Elmer Cetus DNA Thermal Cycler. 10 µl of cDNA obtained from the procedures described above, was added to the PCR reaction mixes which contained a buffer, deoxynucleotides and primers in the recommended concentrations (13). The primers were the adaptor (13; 5'-GACTCGAGTCGACATCG3') and primer #1 (Fig 1; coordinates 1973–1954; 5'-ATCATTGATCATGCTA-CATC3'). Primer #1 was checked for the fidelity of its priming by carrying out a dideoxy sequencing reaction on the plasmid pDSV1 (2).

The conditions for the PCR reactions were 40 cycles of the following sequence; denaturation at 94°C for 40 seconds, a linear rate of cooling from 94°C to 52°C over 2 minutes, annealing temperature at 52°C for 1 second, a linear rate of warming from 52°C to 72°C over 2 minutes and extension temperature of 72°C for 3 minutes. At the end of 40 cycles the reactions were incubated at 72°C for 20 minutes and then stored at 5°C.

Southern hybridisation and preparation of probes

These methods have been described previously (2). Typically, 10% of the PCR product from one reaction was subjected to electrophoresis, blotted and probed with DSV genome length DNA (2).

For hybridisation using the intron specific probe, 40 pmols primer #2 (Fig 1; coordinates 1706–1725; 5'-TCCTCGCGGG-TATAACAGGT3') was labelled using γ[³²P]ATP and polynucleotide kinase (17). Blots were prehybridised for 4 hours and hybridised for 18 hours in 6×SSC (90mM sodium citrate, 0.9M sodium chloride), 10×Denhardt's solution (0.2%, (w/v) Ficoll, 0.2% (w/v) BSA, 0.2% (w/v) polyvinylpyrrolidone) at room temperature. Blots were washed with 6×SSC at 37°C (5×10 minutes).

Molecular cloning

Standard procedures were employed for molecular cloning and transformation of *Escherichia coli* (*E.coli*) (17).

Cloning of PCR amplified cDNAs

The products of one PCR reaction were cut with appropriate restriction enzymes (see Results), and the DNA electro-eluted from agarose gels. The DNA was ligated into compatible, dephosphorylated sites of M13mp18 (19). Positive plaques were identified by hybridisation to a genome length DSV specific probe (19). Dideoxy sequencing (20) was carried out as described previously (21).

Construction of pDSV13

First, a chimaeric luciferase gene (35S-*luc*) was constructed as follows: pKW101 (a kind gift from Prof. D. Helinski, University of California) which contained a nearly complete cDNA of the firefly luciferase gene (22) was used as the source of the *luc* coding sequence. The cDNA was missing 18bp 5' to the first base of the methionine initiation codon (23). A complete coding sequence was obtained by ligating a 40bp SstI-HaeIII fragment created from two oligonucleotides (5'-CAGGCCTATGGAAGACGCCAAAAACATAAAGAAAGG-3' 3'TCGAGTCCGG-ATACCTTCTGCGGTTTTTGTATTTCTTTCC-5') and a 531bp HaeIII-EcoRI (coordinates 29–588) from pKW101 into the SstI and EcoRI sites of pUC19 (18), creating pUC19lucSon. The SstI-HindIII sites in the polylinker of pUC19lucSon were removed by cutting with the restriction enzymes, rendering them blunt using T4 DNA polymerase I and religating, creating plasmid pUC19lucSonpl-. A set of restriction sites from pUC18 was introduced at the 3' end of the partial *luc* sequence by inserting the AatII-EcoRI fragment from pUC18 (2617–447; 18) into the same sites of pUC19lucSonpl-, to create pUC19lucSpl. Finally, the full *luc* coding sequence was restored by inserting the 1203bp EcoRI fragment from pKW101 (coordinates 588–1791) into pUC19lucSpl, to create pJIT27. Therefore, the *luc* coding sequence of pJIT27 has a pUC18 polylinker at its 3' end and an unique StuI site immediately adjacent to the ATG initiation codon.

To create a cauliflower mosaic virus (CaMV) 35S promoter (35S) –*luc* fusion, the *luc* coding sequence was recovered as a StuI-BamHI fragment from pJIT27 and ligated into the HindIII (filled in with T4 DNA polymerase) BamHI sites of pJIT30 which is a pUC based 35S expression cassette (16). This plasmid was called pJIT53.

Finally, pDSV13 was assembled. The starting plasmid was pDS2 which is a tandem dimeric clone of the DSV genome in

pBIN19 (7,24). The 35S-*luc* gene was inserted as a 2.6kb KpnI(T4 DNA polymerase treated)-BglII fragment from pJIT53, into the unique SalI(T4 polymerase treated) and BamHI sites respectively of the pBIN19 polylinker of pDS2. This ensured that the 35S-*luc* gene was inserted on the left border side of the pDS2 polylinker with the direction of transcription of the chimaeric *luc* gene away from the DSV DNA. Therefore, the transfer of the 5.4kb DSV dimer in the T-DNA of pDSV13 can be screened for by selecting kanamycin resistant, luciferase positive transformants.

Tobacco transformation and screening

pDSV13 was mobilised into *Agrobacterium tumefaciens* strain LBA4404 (25) using a triparental mating technique (24). Transformation of tobacco with LBA4404 has been described previously (16).

The luciferase assay was carried out on plantlets which had rooted on medium containing 50mg l⁻¹ kanamycin. Approximately 5mm² piece of leaf was removed from each plantlet and placed in a sterile preweighed plastic scintillation vial containing 200μl of luciferin solution (0.1M citrate buffer [pH5], 1mM D-(-)-luciferin [Sigma]). The vial was reweighed then incubated at room temperature for 1 hour. Light emission was determined with a LKB Rack Beta scintillation counter using the ³H channel set in non-coincidence mode. Light emitted by the plant material was expressed as cpm/mg fresh weight. Untransformed tobacco or those transformed with pBIN19 alone gave background readings of no greater than 180000 cpm/mg. Plantlets were deemed positive if they gave values > 250000 cpm/mg.

DNA was prepared from selected kanamycin resistant, luciferase positive plants, blotted and probed with high specific activity ³²P-labelled DNA as previously described (16). 10μg of tobacco DNA was cut with HindIII, blotted and probed with full length DSV DNA. Copy number of the integrated sequences was determined by densitometry of the hybridised 2.7kb DSV HindIII fragment in comparison with 10, 4, 2 and 1 copy per 2n genome equivalents of linear DSV DNA from pDSV1 (2).

RESULTS

Restriction mapping of DSV specific PCR amplified cDNAs. cDNA prepared from oligo-dT selected RNA, isolated from DSV infected leaf tissue, was PCR amplified using primer #1 and the adaptor primer (Fig 1 and Materials and Methods). Two bands of differing intensities were detected on Southern blots probed with genome length DSV DNA; a major band (a) of ca. 0.9kb in size and a minor band (b) of 0.8kb (Fig 2A, track 3). A further band (c) which almost co-migrated with band a was detected during restriction analysis (Fig 2A; track 4). No DSV specific bands were detected in samples of PCR amplified cDNAs prepared from either healthy *Digitaria* tissue or from RNAase treated poly(A)+ RNA from infected tissue (Fig 2A & B, Tracks 1 and 2), showing that the primers did not recognise any host specific cDNAs which subsequently hybridised to DSV specific probes, nor were the bands caused by amplification of contaminating DSV DNA present in the RNA preparations.

A summary of the restriction enzymes used, the expected sizes of bands for cDNAs derived from the spliced RNA (RNA 4-; 2) and the most abundant unspliced RNA (RNA 2-; 2) is shown in Fig 1.

B=BclI, Bg=BglII, H=HindIII, S=StuI, Sc=ScaI.

Figure 1. The DSV genome between coordinates 2701 and 1000 (6). The major spliced and unspliced transcripts encoded by this region and their putative polypeptides are from Accotto et al (2). +1 refers to the position of the 5' ends of RNAs 2- and 4- and the shaded area indicates the position of the intron (2). The priming sites and the direction of second strand synthesis of the oligonucleotides used in this work (#1 and #2) are shown. Below the coordinates are the positions of the restriction sites used in this study and the predicted sizes of PCR amplified cDNAs and restriction fragments corresponding to RNAs 2- and 4-. The cDNAs were amplified between primer #1 and the adaptor primer (SalI/XhoI/ClaI; see Materials and Methods).

StuI, ScaI and HindIII all have single sites in the C1 or C2 ORFs (Fig 1; 6) which allowed accurate mapping of the cDNAs onto the DSV genome. Of particular use was the HindIII site which is located in the intron (Fig 1). BglII was used to more precisely map the 3' ends of the cDNAs (Fig 1).

Fig 2A (tracks 4-7) shows a Southern blot of the restriction enzyme digested, DSV specific bands a, b and c (Fig 2A, track 3). From this analysis it was possible to determine the RNAs which corresponded to bands a and b.

The major band a (Fig 2A, track 3) was cDNA derived from unspliced RNA species amplified between coordinate 1973 and their 3' end at about 1060, since the restriction digests generated a set of bands which matched the predictions (Fig 1). Furthermore, band a and predicted subsets of its restriction products (Fig 1) hybridised to an intron specific probe (oligonucleotide #2; Fig 1 and Fig 2B tracks 4-7).

Similarly, band b (Fig 2A, track 3) was established to be the spliced version of band a, because it also generated a predicted set of restriction fragments (Fig 1 and Fig 2A, tracks 4-7). In addition, neither band b nor any of its restriction fragments hybridised to the intron specific probe (Fig 2B, tracks 3-7). Of most use in identifying the spliced cDNA was the StuI digest (Figs 2A & B; track 5) which generated a readily identifiable 285bp band.

Band c did not fit any prediction. This band only became apparent during the restriction analysis (Fig 2A, tracks 4-7). The most distinctive difference was the StuI digest which generated an approximately 460bp band (Fig 2A, track 5). This

Figure 2. Southern blot of PCR amplified DSV specific cDNAs, synthesised from the equivalent of 1ng of poly(A) RNA, probed with genome length DSV DNA (panel A) or with the intron specific primer #2 (panel B and Fig 1). In each panel the tracks are as follows: cDNA prepared from healthy *Digitaria* tissue (track 1) or from RNAase treated RNA prepared from DSV infected *Digitaria* tissue (track 2) and incubated in the 3'RACE PCR reactions. Tracks 3–7 are the 3'RACE PCR amplified cDNAs prepared from DSV infected *Digitaria* tissue and either uncut (Track 3) or cut with HindIII (track 4), StuI (track 4), ScaI (track 6) and StuI/BglII (track 7). Size markers in panel A are 5ng pDSV1 (2) cut with StuI/BglII (track 8), ScaI/BglII (track 9) and HindIII (track 10). Their sizes are shown to the right of the panel. In panel B, the same size markers were used as in A, but only the bands predicted to hybridise to the intron specific probe appeared, demonstrating its specificity. Bands a, b and c are the species generated from the PCR reactions (track 3) and their restriction fragments (tracks 4–7) and are discussed in the text. The smeared cDNA band representing the minor population of heterogeneous RNAs (see text) are indicated by the bracket.

band also hybridised to the intron specific probe (Fig 2B; track 5). The other restriction digests were more equivocal, since band c was always obscured by band a products (eg. Fig 2 A & B, track 6). Although not conclusive at this stage, band c was thought not to contain a HindIII site (Fig 2 A & B, track 4), despite hybridising to the intron specific probe (Fig 2B, tracks 3–7). Fragment c was not generated by alternative priming of oligonucleotide #1 further upstream of its intended site since the HindIII digest (Fig 2 A & B, track 4) would have generated a band of approximately 310bp.

To further analyse band c, StuI/ScaI and StuI/ScaI/HindIII digests were carried out on the amplified cDNAs, the Southern blot of which is shown in Fig 3. The cDNAs corresponding to the unspliced (band a) and spliced (band b) gave rise to restriction fragments of a predicted size (Fig 1). Band c generated a 390bp StuI/ScaI band (Fig 3, track 4) which did not contain a HindIII site (Fig 3, track 5). Since no band was greater in size than that generated by the 5' end of the cDNA (1973) to the ScaI site at 1878, we concluded that band c was an unspliced cDNA containing an insertion of 90bp of DNA in the HindIII site (at coordinate 1743).

In addition to the bands a,b and c, two DSV specific smears could be detected, centred around 0.67kb and 0.4kb (Fig 2 A & B; track 3). The smear shifted mobility on digestion with restriction enzymes (eg. Fig 2 A & B, tracks 4 and 6). Assuming that the 5' end of these cDNAs were delineated by primer #1 (Fig 1), then the smear could have corresponded to a minor population of DSV specific RNAs which have highly heterogeneous 3' ends or alternative splice sites.

It is worth noting that no amplified cDNAs corresponding to RNAs 3– and 5– (2) were detected. For example, cDNAs corresponding to RNAs 3– and 5– would have generated 543bp and 750bp HindIII bands respectively.

Figure 3. Further analysis of band c. Southern blot of PCR amplified cDNA from poly(A) RNA prepared from DSV infected *Digitaria* tissue probed with genome length DSV DNA (see legend Fig 2), digested with ScaI (track 3), StuI/ScaI (track 4) and StuI/ScaI/HindIII (track 5). Size markers (left of the panel) are pDSV1 (5ng) cut with StuI/BglII (track 1; 0.38kb), StuI/ScaI (track 2; 0.29kb) and StuI/HindIII (track 6; 0.16kb). Bands a,b and c are as in Fig 2.

Molecular cloning and sequence analysis of PCR amplified cDNAs

PCR amplified DSV cDNA was cut with BclI (1964; Fig 1) and SalI (in adaptor; Fig 1 and Materials and Methods) and ligated into the BamHI and SalI sites of M13mp18 (see Materials and Methods; 18).

The results of the sequence analysis of the positive clones recovered from this first cloning strategy are shown in table 1. Typically only ca. 10% of clear (lac–) plaques contained DSV specific inserts. Only cDNA clones containing a polyA tail were included in the analysis. The size of the poly(A) tail ranged from 15–52 bases with an average size of 22 bases. Those without poly(A) tails accounted for a further 11% of the clones. Spliced cDNA clones were initially identified by absence of the HindIII site at coordinate 1743 (Fig 1) and the failure of the intron specific oligonucleotide #2 (Fig 1) to prime sequencing reactions from the M13 template ssDNA (data not shown). The StuI

Table 1. Summary of the analysis of DSV complementary sense 3' cDNA clones. The coordinates are from the virion sense strand of the DSV sequence (6). The consensus polyadenylation sequence of animal genes (AATAAA) is from Proudfoot and Brownlee (29). Putative alternative polyadenylation signals from three plant genes are from Dean et al (30).

3' end of cDNA (coordinate)	Nearest poly(A) signal	Number of clones	Number spliced
1063	AATAAA;1083	24	6
1067	AATAAT;1086	3	0
1087	ATTAAT;1201	2	0
1104	ATTAAT;1201	1	1
1439	AATAAT;1564	1	0
1450	AATAAT;1564	1	0
1552	ATAATT;1578	1	0
		33	7

(1587)–ScaI (1878) fragment (Fig 1) from each candidate was subcloned into the SmaI site of M13mp18 and sequenced in both strands. In all 7 clones (table 1) the splice junction was located between coordinates 1666 and 1759 as predicted previously (2). The number of spliced cDNA clones obtained (21%) was consistent with the relative abundance of spliced to unspliced RNAs (Fig 2A and 2).

Based on the size of the insert, no clones were recovered which corresponded to band c, although approximately the same number could be expected as for the spliced cDNAs (band b), based on the near equal intensities of bands b and c in the Southern blots (Fig 2A).

In a further attempt to clone the region of band c around the HindIII site at 1743, the 390bp StuI/ScaI fragment of band c (Fig 3) was eluted from an agarose gel and a fraction of it again subjected to electrophoresis, Southern blotted and shown to be ca. 90% 390bp fragment. The remaining 10% was due to equal amounts of contaminating StuI/ScaI fragments derived from bands a and b (data not shown). The remainder of the eluted fragment was ligated into the SmaI site of M13mp18 after digestion with HindIII to reduce still further the amount of contaminating fragment derived from band a. 40 positive plaques were obtained, of which 34 were the StuI/ScaI fragment from the spliced band b and the remainder were equivalent to band a. No recombinant phage containing 390bp inserts equivalent to the StuI/ScaI fragment of band c (Fig 3) were recovered. We concluded that the insert in band c, in some way prevented this fragment from being cloned. Attempts at direct sequence analysis of band c were precluded by the presence of non-specific DNA present in the preparations. This DNA did not hybridize to DSV sequences (data not shown).

Complementary sense transcription of DSV DNA in transgenic tobacco

Tobacco was transformed with pDSV13. 20/42 kanamycin resistant plantlets, derived from different leaf explants were luciferase positive (see Materials and Methods). The 6 independent transformants with the highest level of luciferase activity were selected for further study. The six luciferase positive tobacco plants each contained between 1 and 4 copies of the DSV dimer DNA per 2n genome (data not shown). In the same blot, no replicating forms of DSV DNA were detected (data not shown). This is in contrast to the component A of TGMV, where high copy number episomal forms are generated from the integrated sequences in transgenic petunia and tobacco (26, 27).

Total RNA was prepared from these 6 DSV transformed

Figure 4. Southern blot of DSV complementary sense specific, PCR amplified cDNAs from six independent tobacco plants transformed with pDSV13 (tracks 5–16; see Materials and Methods) and DSV infected *Digitaria* tissue (tracks 1–3). Amplified cDNAs from the equivalent of 100ng of total RNA were used per track. Track 4 is cDNA prepared from untransformed tobacco and incubated in the same PCR reactions. Amplified cDNAs derived from *Digitaria* or tobacco were digested with HindIII (track 2 and tracks 5–10 respectively) and StuI (track 3 and tracks 11–16 respectively). Track 1 is uncut *Digitaria* derived DSV specific cDNA. Bands a, b and c and the bracket are as in Fig 2.

Figure 5. Virus specific transcripts mapped onto the DSV genome. The transcripts are numbered according to Accotto et al, (2). The positions of the ORFs, consensus promoter (◊) and polyadenylation signals (◆) are from Donson et al (6). The position of the intron (see text) is indicated by the gap in RNA 4– and the black arc on the circle.

tobacco plants and using the 3'RACE procedure as described, cDNA corresponding to DSV complementary sense RNA was readily detected. HindIII and StuI digests of the amplified cDNA were blotted and probed with genome length DSV DNA (Fig 4). The bands seen after the HindIII digestion showed that the major cDNA species were the same as their counterparts from DSV infected *Digitaria* (Fig 4, track 2 and 5–10). However, the StuI digested cDNA showed that only the bands corresponding to unspliced RNA (band a) were present (Fig 4, tracks 11–16). Neither the 285bp StuI band, diagnostic for spliced cDNAs, nor the 390bp band derived from band c could be detected (Fig 4, tracks 11–16). cDNA prepared from RNA isolated from untransformed tobacco gave no DSV specific bands (Fig 4, track 4). No DSV specific smear could be detected in the DSV cDNA prepared from tobacco (Fig 4; tracks 5–10) as was observed with DSV cDNAs prepared from *Digitaria* (Fig 2 and Fig 4; tracks 1 and 2). There was variation in the amount of DSV cDNA

detected between individual transformants (Fig 4, tracks 5–16), which did not correlate with the relative levels of luciferase activity or the copy number of the integrated DSV DNA. For example, the tobacco with the lowest level of DSV cDNA (Fig 4, tracks 7 and 13) had the second highest luciferase activity. A lack of correlation between expression and the copy number of integrated DNA has been described previously (16,28).

DISCUSSION

The 3'RACE procedure (13) offers considerable advantages over the nuclease S1 mapping procedures (12) employed to study DSV complementary sense transcription (2). First, PCR amplification makes the analysis much more sensitive. For example, to visualise DNA fragments which protected RNA 4– from nuclease S1 digestion, using 2 µg of poly(A)+ RNA, autoradiographs required exposure for 2 weeks. In contrast, the autoradiographs shown in Figs 2–4, were exposed overnight and were equivalent to 1 ng of poly(A)+ RNA. Second, restriction analysis of the amplified cDNAs allows a view of the transcription of a gene or genes and the processing of its RNA products to be obtained quickly. Third, the amplified cDNA can be cloned and sequenced.

The 3'RACE procedure is better for studying geminivirus transcripts than other PCR procedures which employ two specific primers (11). This is because of the possibility of contamination of the RNA preparations with viral DNA (8,10). The presence of a poly(A) tail on the 3' end of a clone generated from 3'RACE amplification reaction confirms that it was derived from RNA and not from contaminating DNA. This is particularly important in defining cDNAs which contain introns.

The positions of DSV encoded transcripts on the DSV genome are shown in Fig 5 and are the combined results of this work and that of Accotto et al (2). The PCR 3'RACE procedure has allowed spliced DSV specific RNAs to be characterised, confirming the predictions made from the S1 nuclease mapping and homology to consensus donor and acceptor sites (2) and has shown that unspliced steady state RNAs containing a 92bp intron are encoded by the same region of the DSV genome (Fig 2, Table 1 and 2). The 3' ends of these RNAs have been mapped precisely, in the majority of cases to coordinate 1063 (Table 1), downstream from a previously identified putative consensus polyadenylation sequence (6) at coordinate 1083. Using this technique, we have been able to demonstrate formally that DSV does encode polyadenylated RNA species.

Two RNAs have now been removed from the original transcript map (2), RNA 5– was assigned previously on the basis of an internal S1 nuclease sensitive site detected during 2-D electrophoretic analysis of S1 nuclease protected DSV DNA fragments, although the possibility that it might be an artifact due to the presence of a potential hairpin loop at the position of the internal S1 nuclease cleavage site was recognised at the time (2). No splice site in the position mapped for RNA 5– could be detected from the restriction analysis of the PCR amplified DSV specific cDNAs in this work (Fig 2), thus confirming RNA 5– as an artifact. Likewise, the 3' end of RNA 3– was mapped to position 1200 (2) using S1 nuclease mapping procedures, but no equivalent 3' RACE product could be detected (Fig 2). Artifactual 5' and 3' ends can be generated by A-T rich regions which are able to separate into single stranded loops during the protection assay (8). Such a region is present from coordinates 1194–1207 on the DSV genome. These sequences would not have primed polydT during cDNA synthesis, because the region is predominantly T rich in the complementary strand.

Band c was interpreted as cDNA containing an insertion of 90bp at or near the 5' region of the intron (Fig 3). However, restriction fragments could not be cloned and this species was not present in the tobacco derived DSV cDNAs (Fig 4). The proximity of the putative insert to the donor sequences of the intron and its absence in tobacco derived cDNAs leads us to suggest that band c is associated with the presence of spliced cDNAs (band b). Band c could be an artifact of the cDNA synthesis and/or the PCR reactions, although to our knowledge no such PCR mediated insertion of DNA into an amplified cDNA has been reported previously. Alternatively, band c may be the amplified cDNA product of a rare RNA species. If so, such a species would not have been detected by S1 nuclease protection since any fragments generated would have co-migrated with those from spliced RNA 4– (Fig 5; 2). However, these considerations offer no reason for our failure to clone band c. 3'RACE can sometimes generate single-stranded DNA (13), but this was not produced since all types of DSV specific cDNAs could be cut by restriction enzymes (Figs 2 and 3).

The PCR analysis has revealed an hitherto undetected population of DSV encoded RNAs with extremely heterogeneous 3' ends (Fig 2). We do not consider these cDNAs to be artifacts of the cDNA synthesis or PCR reactions since no such smear could be detected from the tobacco derived cDNAs (Fig 4) and examples of cDNAs with ends mapping to this area have been cloned (table 1). The putative polyadenylation signals which are presumed to give rise to these RNAs (Table 1) are different from the animal consensus sequence (AATAAA; 29), but have been identified within the 3' untranslated sequences of the *rbcS* and *Cab* genes of petunia, and the *bz* gene of maize. In these genes, the putative polyadenylation signals are located 16–130 nucleotides from the 3' end of their RNA (30). The absence of these minor RNA species in transformed tobacco suggests that they are associated only with replicating DSV DNA. The significance of these minor RNA species is difficult to assess, but their extremely low abundance and heterogeneous nature suggest that they could be minor 3' processing events caused in some way by host transcription machinery having to interact with actively replicating high copy number viral DNA.

The detection of cDNAs mapping to the C1 and C2 ORFs in tobacco plants transformed with a tandem dimer of DSV DNA shows that the complementary sense promoter (2) is active in tobacco and that the 3' end of the RNA is correctly processed (Fig 4). However, the same mapping of the cDNAs revealed that no splicing of the RNA occurred (Fig 4). The failure of the RNA to splice and consequent failure to translate C2 sequences may explain why DSV does not replicate in tobacco since the C2 ORF has been shown to be absolutely required for ds DNA replication in monocot infecting geminiviruses (11).

There are a few examples of monocot genes being transferred to dicots and their expression studied. A maize zein gene and a wheat *Cab* gene have been successfully expressed in sunflower callus, tobacco and petunia respectively, although neither of these genes contain introns (31,32). The maize transposable elements *Activator* (*Ac*) and *Enhancer* (*En/Spm*) have been shown to function in tobacco so presumably their RNAs are synthesised and processed correctly (33,34). *Ac* contains 4 introns and *En* undergoes complex splicing of its transcripts (35,36). In contrast, the promoters of the maize *Adh* and wheat *rbcS* genes do not function in tobacco (37,38). However, when the genes are transcribed from a CaMV 35S promoter their introns are spliced inefficiently and their polyadenylation signals function aberrantly (38).

The DSV intron possesses donor and acceptor junctions which match consensus sequences (2) and its size (92bp) is typical of plant introns (39). The A+U content of the DSV intron is 55.4%, a value comparable to introns from other monocot genes (40), but is low compared with dicot introns and may explain why the DSV intron failed to splice in tobacco (39,40).

It should be emphasised that in addition to being non-functional in tobacco the DSV intron is only partly functional in *Digitaria*. This is analogous to the partial splicing observed for the wheat *rbcS* and maize *Adh1* introns in tobacco (38) and the behaviour of artificial introns in transiently expressing tobacco and maize protoplasts (39). In contrast, in maize and wheat no unspliced mature RNAs of the *Adh1* and *rbcS* genes have been reported to accumulate. Thus, DSV may possess an inefficiently spliced intron in order to ensure the synthesis of both C1 and C1:C2 products from RNAs 2- and 4- respectively (Fig 1 and 5: 2).

It is known that small changes in the sequences of an intron can produce differences in splicing efficiency between monocot and dicot cells (39). Therefore, only a small region of the DSV intron may be responsible for its inefficient splicing in *Digitaria* and failure to function in transgenic tobacco.

Alternatively, the splicing process in DSV may be more controlled. The failure of the DSV complementary sense RNA to splice in tobacco and its partial splicing in *Digitaria* may be due to the absence of a component of the splicing machinery in dicots which in monocots is expressed either in limiting amounts or is regulated temporally or in a tissue specific manner. In this respect, the report that there are multiple forms of U2 and U4 small nuclear (sn) RNAs from monocots compared with a more homogeneous population of snRNAs in dicots and that a U1-like snRNA is differentially expressed in wheat (41) may be significant.

ACKNOWLEDGEMENTS

We would like to thank Helen Reynolds for preparing RNA from tobacco, Prof. D. R. Helinski for pKW101 and for sequence data before publication, Dr. Noel Ellis for useful discussion during the course of this work and Dr. Gary Creissen for critical reading of the manuscript. This work was supported by a grant-in-aid from the AFRC to the John Innes Institute and by the DTI Gene Took Kit Consortium. This work was carried out under MAFF licence number PHF 1185/89 (89).

REFERENCES

- Howarth, A.J. and Vandemark, G.J. (1989) *J. Gen. Virol.*, 70, 2717-2727.
- Accotto, G.P., Donson, J. and Mullineaux, P.M. (1989) *EMBO J.*, 8, 1033-1039.
- Stanley, J. (1985) *Adv. Virus Res.*, 30, 139-177.
- Abouzid, A.M., Frischmuth, T. and Jeske, H. (1988) *Mol. Gen. Genet.* 212, 252-258.
- DePamphilis, M.L. and Wasserman, P.M. (1982) In Kaplan, A.S. (ed.), *Organisation and Replication of viral DNA*. CRC Press, Boca Raton, FL, USA, pp. 38-114.
- Donson, J., Accotto, G.P., Boulton, M.I., Mullineaux, P.M. and Davies, J.W. (1987) *Virology*, 161, 160-169.
- Donson, J., Gunn, H.V., Woolston, C.J., Pinner, M.S., Boulton, M.I., Mullineaux, P.M. and Davies, J.W. (1988) *Virology*, 162, 248-250.
- Townsend, R., Stanley, J., Curson, S.J. and Short, M.N. (1985) *EMBO J.*, 4, 33-38.
- Sunter, G., Gardiner, W.E. and Bisaro, D.M. (1989), *Virology*, 170, 243-250.
- Morris-Krsinich, B.A.M., Mullineaux, P.M., Donson, J., Boulton, M.I., Markham, P.G., Short, M.N. and Davies, J.W. (1985) *Nucleic Acids Res.* 13, 7237-7256.
- Schalk, H.-J., Matzeit, V., Schiller, B., Schell, J. and Gronenborn, B. (1989) *EMBO J.*, 8, 359-364.
- Favaloro, J.M., Treisman, R.H. and Kamen, R.I. (1980) *Methods Enzymol.*, 65, 718-749.
- Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA.* 85, 8998-9002.
- Mullineaux, P.M., Boulton, M.I., Bowyer, P., van der Vlugt, R., Marks, M., Donson, J. and Davies, J.W. (1988) *Plant Mol. Biol.* 11, 57-66.
- Grimsley, N., Hohn, T., Davies, J.W. and Hohn, B. (1987), *Nature* 325, 177-179.
- Guerineau, F., Brooks, L., Meadows, J., Lucy, A., Robinson, C., and Mullineaux, P. (1990), *Plant Mol. Biol.*, 15, 127-136.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene*, 33, 103-119.
- Grunstein, M. and Hogness, D.S. (1975) *Proc. Natl. Acad. Sci.*, 72, 3961-3965.
- Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.H. (1980). *J. Mol. Biol.* 98, 161-178.
- Woolston, C.J., Barker, R., Gunn, H., Boulton, M.I. and Mullineaux, P.M. (1988) *Plant Mol. Biol.* 11, 35-43.
- de Wet, J.R., Wood, K.V. Helinski, D.R. and DeLuca, M. (1985) *Proc. Natl. Acad. Sci. USA*, 82, 7870-7873.
- Ow, D., Wood, K.V., DeLuca, M., de Wet, J.R., Helinski, D.R. and Howell, S.H. (1986) *Science* 234, 856-859.
- Bevan, M. (1984) *Nucl. Acids Res.* 12, 8711-8721.
- Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J. and Schilperoort, R.A. (1983), *Nature*, 303, 179-180.
- Rogers, S.G., Bisaro, D.M., Horsch, R.B., Fraley, R.T., Hoffmann, N.L., Brand, L., Elmer, J.S., and Lloyd, A.M. (1986) *Cell*, 45, 593-600.
- Hayes, R.J. and Buck, K.W. (1989) *Nucl. Acids Res.*, 17, 10213-10222.
- Jones, J.D.G., Dunsmuir, P. and Bedbrook, J. (1985) *EMBO J.*, 4, 2411-4218.
- Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214.
- Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1985) *Nucl. Acids Res.*, 14, 2229-2240.
- Matzke, M.A., Susani, M., Binns, A.N., Lewis, E.D., Rubenstein, I. and Matzke, A.J.M. (1984) *EMBO J.*, 3, 1525-1531.
- Lamppa, G., Nagy, F. and Chua, N.-H. (1985) *Nature* 316, 750-752.
- Baker, B., Schell, J., Lorz, H. and Fedoroff, N. (1986) *Proc. Natl. Acad. Sci. USA*, 83, 4844-4848.
- Pereira, A. and Saedler, H. (1989) *EMBO J.*, 8, 1315-1321.
- Kunze, R., Stochaj, U., Laufs, J. and Starlinger, P. (1987) *EMBO J.*, 6, 1555-1563.
- Gierl, A., Lutticke, S. and Saedler, H. (1988) *EMBO J.*, 7, 4045-4053.
- Ellis, J.G., Llewellyn, D.J., Dennis, E.S. and Peacock, W.J. (1987). *EMBO J.*, 6, 11-16.
- Keith, B. and Chua, N.-H. (1986) *EMBO J.*, 5, 2419-2425.
- Goodall, G.J. and Filipowicz, W. (1990) *Plant Mol. Biol.* 14, 727-733.
- Goodall, G.J. and Filipowicz, W. (1989) *Cell*, 58, 473-483.
- Egeland, D.B., Sturtevant, A.P. and Schuler, M.A. (1989), *The Plant Cell*, 1, 633-643.