The EMBO Journal vol.8 no.4 pp.1033 - 1039, 1989

Mapping of *Digitaria* streak virus transcripts reveals different RNA species from the same transcription unit

Gian Paolo Accotto[1], Jonathan Donson and Philip M.Mullineaux

John Innes Institute, AFRC Institute for Plant Science Research, Colney Lane, Norwich NR4 7UH, UK and [1]Istituto di Fitovirologia Applicata, Via O.Vigliani 104, 10135 Torino, Italy

Communicated by D.A.Hopwood

All, except 18 bp, of the *Digitaria* streak virus (DSV) genome is transcribed from either the virion (two RNA species) or the complementary (up to five RNA species) DNA strand. Detailed mapping of these RNAs has revealed evidence for splicing in one species (RNA 4−), which together with its more abundant unspliced counterpart (RNA 2−) could synthesize both a 30.5 and 41 kd polypeptide from the same transcription unit. This extensive overlapping of spliced and unspliced RNAs could indicate that the initiation and splicing of transcripts is temporally regulated. At least one transcript (RNA 1−) may have a non-translational role. Transcription of the DSV genome shows similarities to some animal DNA viruses, particularly the papovaviruses.

Key words: geminivirus/Gramineae/RNA splicing/single-stranded DNA viruses/transcription

Introduction

The geminiviruses are a group of plant DNA viruses which could be attractive models for the study of gene expression in the Gramineae. The geminiviruses are characterized by twinned quasi-isometric particles, containing genomes of circular single-stranded (ss) DNA. In addition, geminivirus DNA has been isolated from infected plants as a supercoiled double-stranded (ds) form (reviewed in Stanley, 1985) as well as minichromosomes (Abouzid *et al.*, 1988), which by analogy with other DNA viral systems such as the papovaviruses (DePamphilis and Wassermann, 1982) may represent transcriptionally active forms of the virus. The geminiviruses can be divided into two sub-groups (Stanley, 1985), the members of the first sub-group are transmitted by whiteflies, are confined to dicotyledonous hosts, are all serologically related and possess similarily organized bipartite genomes. In contrast, members of the second sub-group are transmitted by different species of leafhoppers, infect either monocotyledonous or dicotyledonous hosts, are serologically distinct and possess monopartite genomes which differ from each other depending upon the host. The subject of this paper, *Digitaria* streak virus (DSV), can replicate in several species of the Gramineae and is a member of the latter sub-group of geminiviruses (Dollet *et al.*, 1986; Donson *et al.*, 1987, 1988; P.G.Markham and H.V.Gunn, unpublished data).

Nucleotide sequence data are now available for several members of each sub-group (Stanley and Gay, 1983; Hamilton *et al.*, 1984; Howell, 1984; Mullineaux *et al.*, 1984; Howarth *et al.*, 1985; MacDowell *et al.*, 1985; Stanley *et al.*, 1986; Andersen *et al.*, 1988; Lazarowitz, 1988; Woolston *et al.*, 1988), including DSV (Donson *et al.*, 1987, 1988) and have revealed differences as well as conserved features in the genome organization of the two sub-groups (MacDowell *et al.*, 1985; Mullineaux *et al.*, 1985; Rogers *et al.*, 1986; Stanley *et al.*, 1986; Donson *et al.*, 1987). Although the nucleotide sequence data have indicated that bidirectional transcription of all the geminivirus genomes may be the norm, confirmation by the mapping of virus encoded transcripts is only available for cassava latent virus (CLV), a member of the first sub-group (Townsend *et al.*, 1985) and maize streak virus (MSV), a member of the second sub-group (Morris-Krsinich *et al.*, 1985). In the case of MSV, detailed analysis of its mRNAs was precluded by the very low level of the complementary sense encoded transcript(s) (Morris-Krsinich *et al.*, 1985).

In this paper a more extensive analysis of the transcripts of a geminivirus is presented. Our study has indicated a more complex pattern of transcription of the geminivirus genome than had hitherto been suspected. Differential splicing of at least one of the virus encoded transcripts occurs, and we propose that regulation of gene expression could be achieved by modulating the maturation of these RNAs. Our data could indicate that temporal control of the initiation of transcription also happens. These observations and the possibility that more than one type of protein may be synthesized from

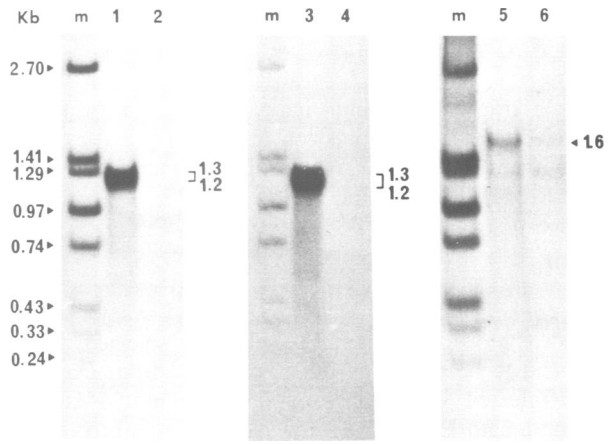

Fig. 1. Northern blot of 0.2 μg (**tracks 1−4**) and 2 μg (**tracks 5** and **6**) of poly(A)$^+$ RNA prepared from infected (**tracks 1, 3** and **5**) and healthy (**tracks 2, 4** and **6**) *Digitaria* plants. The probe was a full length, oligo-primed ^{32}P-labelled DSV *Hin*dIII DNA fragment (**tracks 1** and **2**). For sense specific hybridization (Morris-Krsinich *et al.*, 1985) the blots were first probed with unlabelled full length DSV M13 mp9 *Hin*dIII clones of complementary sense (**tracks 3** and **4**) or virion sense (**tracks 5** and **6**) ssDNA followed by a second probe of oligo-primed, ^{32}P-labelled M13 mp19 RF DNA. The positions of the bands are marked by arrows (◀) to the right of the track. The markers (m) were mixtures of RF form of the DSV M13 mp9 *Hin*dIII clone restricted with *Hin*dIII, *Pst*I−*Hin*dIII and *Kpn*I−*Hin*dIII.

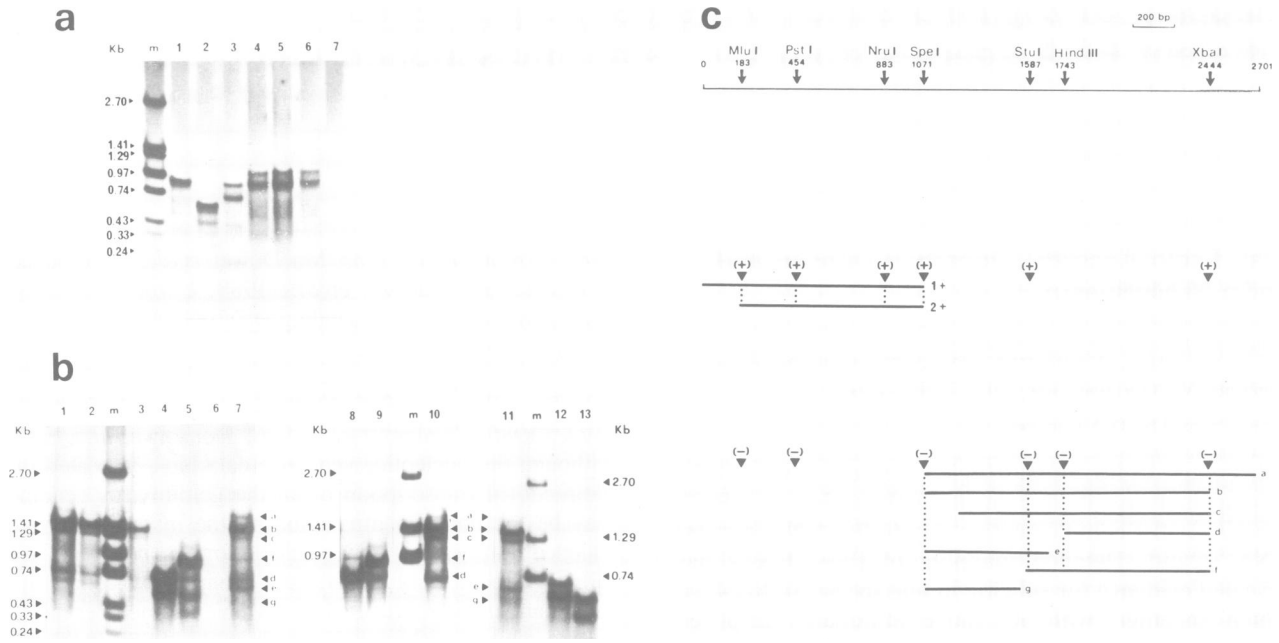

Fig. 2. (**a**) Southern blot of S1 nuclease resistant fragments of complementary sense (−) M13 ssDNA clones [which gives a (+) sense sequence ladder; shown in c] protected by virus specific RNAs. **Track 7** shows the results of a S1 nuclease protection of complementary ssDNA of a full length *Xba*I clone of DSV in M13 mp18 using poly(A)$^+$ RNA prepared from healthy plants. The markers (**track m**) are the same as in Figure 1. (**b**) (**i**) Southern blot of S1 nuclease resistant fragments of virion (+) sense M13 ssDNA clones [which gives a (−) sense sequence ladder as shown in c] protected by virus specific RNAs. **Track 6** shows the results of an S1 nuclease protection of a full length *Pst*I clone of DSV DNA in M13 mp9 using poly(A)$^+$ RNA prepared from healthy plants. The S1 nuclease resistant fragments generated by ssDNA clones (−) *Mlu*I (**track 1**), (−) *Pst*I (**track 2**), and (−) *Spe*I (**track 3**) are labelled a−g and are shown on the right of the panel. The marker DNA fragments (**track m**) are the same as in Figure 1. (**ii**) Southern blots of S1 nuclease resistant fragments of virion sense M13 ssDNA clones (−) *Pst*I (**tracks 10** and **11**), (−) *Hin*dIII (**tracks 8** and **12**) and (−) *Stu*I (**tracks 9** and **11**) protected by virus specific RNAs. Probes used were oligo-primed, ^{32}P-labelled DNA fragments generated by digestion of pDSV1 with *Hin*dIII−*Xba*I (co-ordinates 1743−2444; **tracks 8−10**) and *Spe*I−*Stu*I (co-ordinates 1071−1587; **tracks 11−13**). The S1 nuclease resistant fragments a−g which hybridized to each of these probes are indicated by the arrows in the centre of the two right panels (◀). The mixture of DNA fragments was the same as in Figure 1, but only some were detected by each probe and are indicated by the arrows to the right and left of the panel. (**c**) Protected fragments of the M13 ssDNA clones aligned with the DSV genome. The restriction sites used to construct the M13 clones are marked on the lines representing the length of the genome. The numbers marked below the restriction sites are their co-ordinates determined from the nucleotide sequence of DSV DNA (Donson *et al.*, 1987). The 1.3 and 1.2 kb virion sense RNAs are termed 1+ and 2+ respectively. The thick black lines represent the positions of protected DNA fragments generated from M13 ssDNA clones with cloning sites outside the transcribed sequences. Vertical dotted lines represent the S1 nuclease cleavage sites of RNA−DNA hybrids interrupted by M13 DNA at the cloning junctions.

a single transcription unit indicate a similarity to the organization of gene expression of the papovaviruses and adenoviruses of animals (DePamphilis and Wasserman, 1982; Broker, 1984; Flint, 1986).

Results

Virus-specific poly(A)$^+$ RNAs

Northern blots of oligo(dT)−cellulose selected RNA, presumed to be poly(A)$^+$ RNA, prepared from *Digitaria* plants infected with DSV and probed with ^{32}P-labelled genome length DSV DNA, revealed a single broad band between ~1.2 and 1.3 kb in size (Figure 1, lane 1). No virus specific transcripts were detected in preparations of poly(A)$^+$ RNA prepared from healthy plants (Figure 1, lane 2).

To determine the polarity of viral transcripts, a 'sandwich' blot (Morris-Krsinich *et al.*, 1985) was carried out. ssDNA forms of both orientations of a clone were used as probes for complementary and virion sense transcripts. The use of these probes revealed that the 1.2−1.3 kb RNA was encoded by the virion sense of the DSV genome (Figure 1, lane 3). Northern blots probed for complementary sense RNAs revealed a band of ~1.6 kb in size (Figure 1, lane 5). Faint bands common to RNAs prepared from both healthy and infected tissue were also present (Figure 1, tracks 5 and 6). We considered these bands to be artefacts of the sandwich blot procedure since long exposures of Northern blots probed for the virion sense RNAs revealed the same bands, while Northern blots probed directly with a full length ds insert did not (data not shown).

The steady-state levels of the RNA encoded by complementary sense DNA was ~1% of that encoded by the virion sense DNA (data not shown).

Mapping of the virion sense transcripts

The position of the virion sense transcripts on DSV DNA was determined using S1 nuclease mapping procedures. Poly(A)$^+$ RNA protected DNA fragments of 1.06 and 0.91 kb when complementary sense ssDNA of full length M13 clones were used, with insertion sites at co-ordinates 1071, 1587 and 2444 on the DSV genome (Figure 2a, lanes 4, 5 and 6). When ssDNA of full length M13 clones with insertion sites at co-ordinates 454 or 883 were used, several fragments of differing sizes were generated (Figure 2a, lanes 1, 2 and 3). No S1 nuclease resistant fragments were

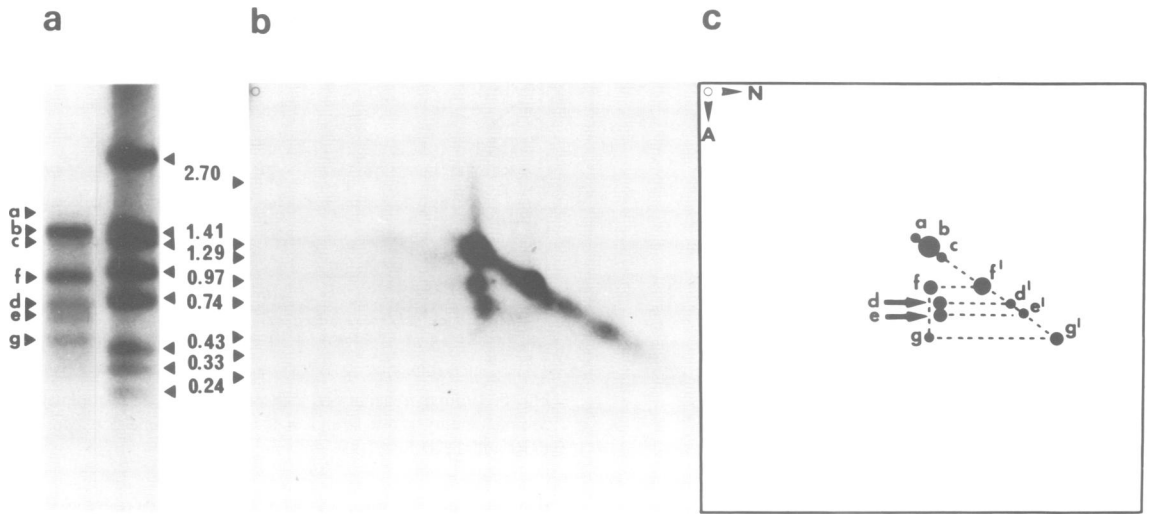

Fig. 3. (**a**) Southern blot of S1 nuclease resistant fragments a−g (see Figure 2b) generated by the protection of virion sense ssDNA of a *Pst*I clone of DSV DNA in M13 mp9 by virus specific RNAs. The bands are indicated by arrows (▶) on the left. (**b**) Southern blot of S1 nuclease resistant fragments using the same M13 clone as in (a), protected by virus specific RNAs, generated by treatment with S1 nuclease at 15°C and subjected to electrophoresis in two dimensions as described by Favaloro *et al.* (1980). The first dimension (horizontal axis) was run in neutral buffer (N; shown in **panel c**) and the second dimension (vertical axis) in alkali buffer (A; shown in **panel c**). (**c**) Interpretation of the blot in **panel b**. Bands a−g correspond to those shown in **panel a** and Figure 2 (**tracks 1, 2** and **3**). Bands d′−f′ are the products generated by complete scission of the RNA−DNA hybrids at the position of their internal S1 nuclease sensitive sites (Favaloro *et al.*, 1980) and correspond to bands d−f. Marker DNA fragments are the same as in Figure 1 and are indicated by arrows between **panels a** and **b**.

generated when poly(A)$^+$ RNA prepared from healthy plants was present in these reactions (Figure 2a, track 7). Thus, the most abundant RNA band detected on Northern blots were two virion sense transcripts of ~1.3 and 1.2 kb in size which had 5′ termini near co-ordinates 0 and 150 respectively, and 3′ termini at approximately co-ordinate 1060 (Figure 2c).

Mapping of the complementary sense transcripts

To locate the positions of the complementary sense encoded transcripts, poly(A)$^+$ RNA was hybridized to virion sense ssDNA of full length M13 clones inserted at co-ordinates 1243, 183, 454, 1071, 1587 and 2444 on the DSV genome and subjected to the S1 nuclease assay. When the protected fragments were analysed by Southern hybridization, a complex series of bands of differing intensities was observed (Figure 2b). When clones with junctions at co-ordinates 183, 454 or 1071 were used (Figure 2b, lanes 1, 2 and 7), seven bands (a−g) were generated. Protection experiments conducted with clones at co-ordinates 1743, 1587 and 2444 gave different sized protected fragments (Figure 2b, lanes 3, 4 and 5). From the information obtained from the protection experiments and the use of specific DSV DNA fragments as probes (Figure 2c), it was possible to map six of the protected DNA fragments onto the DSV genome (Figure 2c). The 5′ and 3′ ends of bands a−f were mapped as follows: a, 2660−1050; b, 2450−1050; c, 2450−1200; d, 2450−1750; e, 1650−1050; and f, 2450−1550. Band g was estimated to be 0.45 kb and was mapped to the region between the *Stu*I and *Spe*I sites, though more accurate mapping was not possible with the data obtained. However, if it is assumed that the 3′ end was co-terminal with most of the other complementary sense transcripts at co-ordinate 1050, then the 5′ end of the transcript corresponding to band g is likely to be around co-ordinate 1500. The intensity of bands f and g was variable between different preparations of RNA, in contrast to bands a−e which remained relatively constant.

Internal S1 nuclease sensitive sites in complementary sense transcripts

RNA−DNA hybrids, formed by hybridization of virus-specific RNAs to virion sense ssDNA of a full length *Pst*I clone of DSV DNA in M13 mp9, were subjected to digestion by S1 nuclease at 15°C. At this temperature complete digestion of internal S1 nuclease sensitive sites is retarded resulting in a nick in the DNA template, but no cleavage of the RNA component in a proportion of the RNA−DNA hybrids (Favaloro *et al.*, 1980). The S1 nuclease resistant RNA−DNA hybrids generated from such an experiment were subjected to electrophoresis in two-dimensional 1.5% (w/v) agarose gels (Favaloro *et al.*, 1980), blotted and probed with genome length DSV DNA. The interpretation of two-dimensional gels has been described in detail by Favaloro *et al.* (1980). Briefly, continuous DNA strands in hybrids between RNAs containing no internal S1 nuclease sensitive sites and their template DNA will have the same relative mobility in the neutral and the alkaline dimensions and therefore will fall along a diagonal line (Figure 3b and c). The component DNA strands of an S1 nuclease digested hybrid formed between a RNA containing a single internal S1 nuclease sensitive site and its template DNA will migrate together in the neutral dimension (held together by the RNA strand), but will resolve in the alkaline dimension as two discrete spots vertically aligned below the diagonal (Figure 3b and c). A line drawn between these spots extrapolates to a point on the diagonal corresponding to the length of the component DNA single strands (Figure 3c). From a comparison of the migration of the DNA fragments in one- and two-dimensional gels (Figure 3a and b), the spots off the diagonal corresponded to bands d and e, and f and g (Figure 3c). The four spots (d and e, co-migrated) below

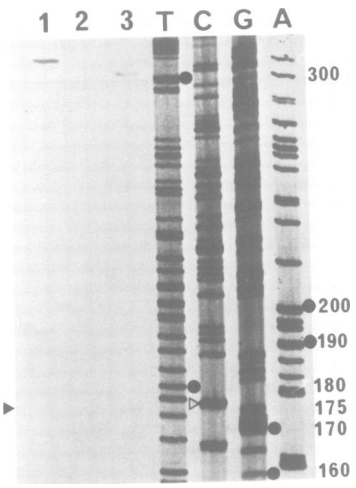

Fig. 4. Location of the internal S1 nuclease sensitive site of the complementary sense transcript corresponding to bands d and e. The high resolution S1 nuclease mapping techniques cited by Morris-Krsinich *et al.* (1985) were used. pDSV1 was digested with *Nco*I, end-labelled using Klenow fragment of DNA polymerase I with [α-^{32}P]dCTP and digested with *Ava*II. After electrophoresis a 309 bp *Nco*I−*Ava*II fragment (co-ordinates 1936−1627) was isolated and used to form hybrids with virus specific RNAs in 3 μg of total poly(A)$^+$ RNA. After digestion of the hybrids with S1 nuclease, protected fragments were subjected to electrophoresis through 6% (w/v) polyacrylamide/urea gels as described previously (Morris-Krsinich *et al.*, 1985). The size of the S1 nuclease resistant fragment present in the treatment with RNA (**track 3**) was determined as described by Donson *et al.* (1984), using a size ladder generated from M13 mp18 employing [α-^{35}S]dATP and universal primer as stated in Materials and methods. Untreated and S1 nuclease treated DNA fragments are shown in **tracks 1** and **2** respectively. The sizes of the bands in the sequencing ladder, in the regions of the protected fragments, are shown to the right of each panel (●). The band in the size ladder used to calculate protected DNA fragments unique to track 3 are indicated (▷).

the diagonal of the gel are consistent with the formation of two RNA−DNA hybrids each with one internal S1 nuclease sensitive region.

Search for splice acceptor and donor sites

One explanation for such internal S1 nuclease sensitive sites is that they represent intron−exon splice junctions (Favaloro *et al.*, 1980). We searched the nucleotide sequence of DSV, in the region of the internal S1 nuclease sensitive sites, for matches to the consensus sequences of Sharp *et al.* (1987) for splice donor and acceptor sites [5′-AG:GTAAGTA \ T-3′ and 5′-(T \ C)nNT \ CAG:G-3′ respectively]. Matches were found at co-ordinates 1759 (5′-AG:GTCAGTG-3′), close to the the 3′ end of band d and 1666 (5′-GATTAC-ACTAG:G-3′), close to the 5′ end of band e, for donor and acceptor sequences respectively. Confirming the sequence analysis, the 3′ end of the exon corresponding to band d was mapped precisely to co-ordinates 1759 (Figure 4). Computer generated splicing and translation of the transcript which generated bands d and e in the protection experiments showed that it would direct the synthesis of a polypeptide which would be the product of a fusion of the 30.5 and 16.9 kd open reading frames (ORFs) (Figure 5, Figure 6, transcript 4−).

Matches were not found to the consensus donor and acceptor sequences in the regions of the 3′ and 5′ ends of

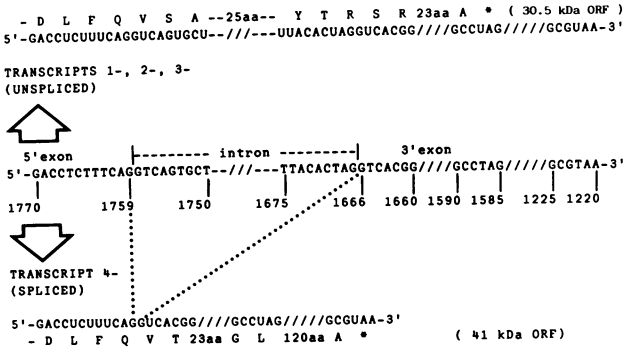

Fig. 5. Computer generated transcription and translation of the complementary sense strand of the DSV genome between co-ordinates 1770 and 1220.

the internal S1 nuclease sensitive site corresponding to bands f and g respectively. A potential hairpin structure has been identified at co-ordinate 1538 (Donson *et al.*, 1987) which could have caused the observed internal S1 nuclease sensitivity. Differences in the levels of bands f and g between experiments may be caused by a variable proportion of the RNA population containing a hairpin structure. No potential secondary structure has been reported in the S1 sensitive region of the transcript corresponding to bands d and e (Donson *et al.*, 1987).

Fine mapping of the 5′ ends of viral transcripts

To confirm that the 5′ co-ordinates of bands a, b, c, d and f were not a consequence of the extensive secondary structure of this region of the DSV genome, the positions on the DSV sequence of the 5′ termini of the viral transcripts were determined.

The 5′ extremities of the two virion sense transcripts were located by a modification of the S1 nuclease protection assay described by Morris-Krsinich *et al.* (1985). In these experiments, the largest S1 nuclease resistant fragments were accompanied by smaller bands migrating up to four bases ahead of it (data not shown). These smaller fragments may either represent some degree of length heterogeneity of the transcripts, or imprecise digestion of the DNA−RNA hybrids by the S1 nuclease. Taking the sizes of the longest S1 nuclease protected fragments, the 5′ termini of the 1.3 and 1.2 kb mRNAs mapped to co-ordinates 2700 and 168 respectively (data not shown and Figure 6).

From the mapping of the 5′ termini of bands b, c, d and f, two sets of S1 nuclease protected fragments could be discerned (groups I and II in Figure 6). The presence of a gap between the two sets, the complexity of the banding pattern for this protection experiment (compared with that presented in Figure 4 and encountered in the mapping of the 5′ ends of the virion sense transcripts) and the probability that the 5′ termini of four transcripts map to this region, suggest that the length heterogeneity observed may reflect the initiation of more than one species of transcript in this region. Taking the sizes of the longest S1 nuclease protected fragments, the 5′ termini of the transcripts which correspond to bands b, c, d and f were calculated to map to co-ordinates 2467 (group I; Figure 6) or 2476 (group II; Figure 6). The 5′ terminus of the transcript corresponding to band a was determined to be at co-ordinate 2680 (Figure 6).

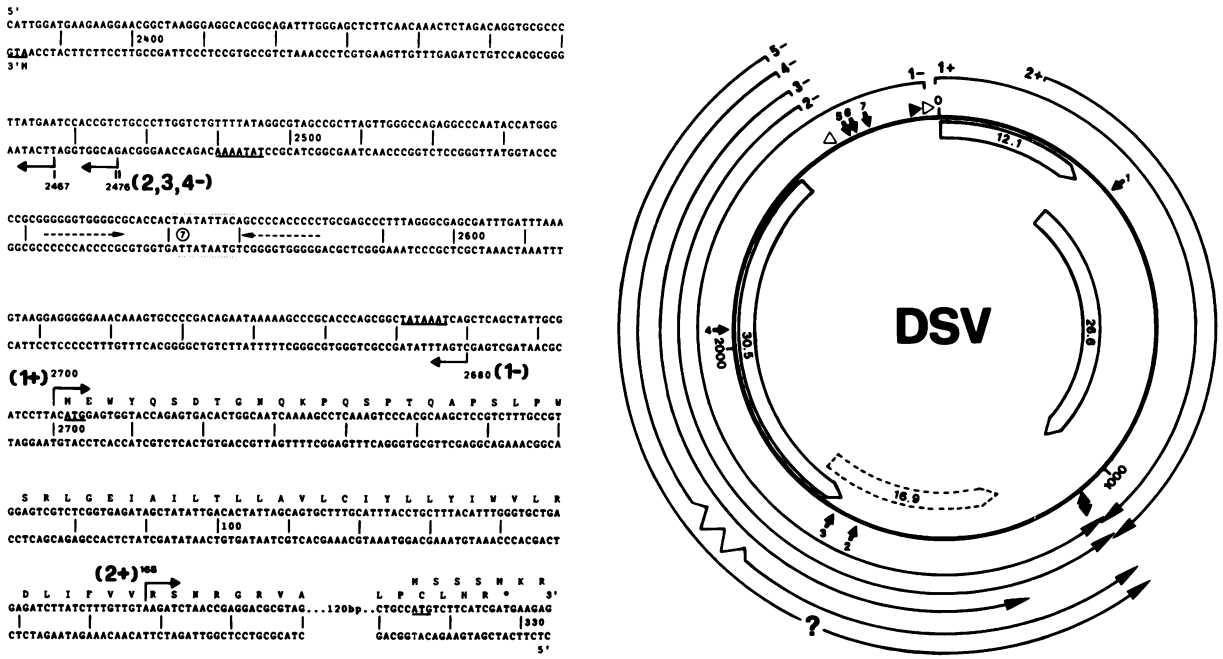

Fig. 6. Virus specific transcripts mapped onto the circular map of DSV genome. The positions of the ORFs, consensus promoters (▷), polyadenylation sequences (▶) and potential hairpin structures ($\Delta G \geq -14$ kcal/mol; arrows numbered 1–7) are from Donson *et al.* (1987). The zig-zag (W) in transcript 4− indicates the approximate position of the intron in the DSV sequence. The question mark (?) in transcript 5− indicates the approximate position of an internal S1 nuclease sensitive site which may have been caused by hairpin 2. Location of the 5′ ends of virus specific transcripts on the DSV sequence (2383–334). The 5′ termini and directions of transcription are marked (⟶▶). Consensus promoter sequences and the initiator codons of the ORFs, identified by Donson *et al.* (1987) and shown on the circular map are underlined. The conserved nonanucleotide sequence in the head of stem–loop 7 (Davies *et al.*, 1987) is marked by a dotted line.

Discussion

The positions of the DSV encoded transcripts in relation to potential transcription regulatory elements and ORFs (Donson *et al.*, 1987) are shown in Figure 6. The 1.3 and 1.2 kb virion sense RNAs are designated 1+ and 2+ respectively. The complementary sense encoded transcripts, corresponding to S1 nuclease resistant bands a, b, c, d and e, f and g, have been termed 1−, 2−, 3−, 4− and 5− respectively. The numbering of the transcripts is not related to their abundance.

The detection of DSV-specific virion and complementary sense RNAs was consistent with bidirectional transcription of the virus genome predicted from its nucleotide sequence (Donson *et al.*, 1987) and by analogy with the transcription of CLV and MSV DNA (Morris-Krsinich *et al.*, 1985; Townsend *et al.*, 1985). However, only 18 bp (co-ordinates 2681–2699) of the DSV genome are not transcribed in at least one strand, in contrast to the non-transcribed region of component 1 of the CLV genome, which is ~278 bp in size (Townsend *et al.*, 1985). In the case of MSV, the complementary sense transcripts were present at very low levels, which prevented their detailed mapping (Morris-Krsinich *et al.*, 1985).

The positions of the 5′ termini of the transcripts in relation to possible promoter sequences, hairpin structure 7 and the translation start site of potential polypeptides is shown in Figure 6. Transcripts 1− and 2+ have no recognizable promoter sequence close to their 5′ termini (<33 nt; Messing *et al.*, 1983), although TATA sequences are recognizable further upstream (Figure 6). All of the mapped RNAs, except transcript 3−, had 3′ termini located near to the consensus polyadenylation sequence AATAAA (co-ordinates 1064 and 1084 in virion and complementary strands respectively) (Figure 6). For transcript 3−, the nearest recognizable sequence is AATAA, at co-ordinate 1394 (Donson *et al.*, 1987).

Both virion sense transcripts could translate the 26.6 kd ORF (Figure 6), which is likely to code for the coat protein of the virus (Donson *et al.*, 1987). In addition, transcript 1+ is likely to be translated to also give the product of the 12.1 kd ORF, a non-structural protein which has been detected for MSV (Mullineaux *et al.*, 1988).

The mapped positions of the DSV specific RNAs reveal that most of their 3′ termini overlap and the sequences of transcripts 2−, 3−, 4− and 5− lie within those of transcript 1−, while those of transcript 2+ lie within 1+ (Figure 6). Overlapping and 3′ co-terminal transcripts are a common feature of the adenoviruses and papovaviruses (DePamphilis and Wasserman, 1982; Broker, 1985; Flint, 1986) as well as CaMV (Covey *et al.*, 1981; Guilley *et al.*, 1982). 5′ co-terminal RNA species are reported for the adenoviruses and papovaviruses (DePamphilis and Wasserman, 1982; Broker, 1984; Flint, 1986) but not for CaMV. Since it is unlikely that proteins involved in the control and initiation of transcription can interact with DNA sequences which are in the process of being transcribed, the detection of overlapping DSV encoded RNA may result from the temporal regulation of transcription of the DSV genome. Virus specific RNAs in this study are derived from a DSV genome of defined sequence (Donson *et al.*, 1987, 1988; and Materials and methods) in contrast to the study of the transcripts of MSV (Morris-Krsinich *et al.*, 1985) which were prepared from plants infected with wild-type virus

known to contain some sequence heterogeneity (Mullineaux *et al.*, 1984). Also, the infection of plants used in this study was asynchronous and therefore it must be assumed that these infected plants contained a mixture of transcripts produced in all stages of the viral life cycle.

The mapping data and sequence searches suggest that there is a 92 bp intron between co-ordinates 1758 and 1667 on the complementary sense DNA of the DSV genome, which is spliced out of a nascent transcript to give RNA 4− (Figures 5 and 6). A second potential splicing event to generate RNA 5− is shown in Figure 6 with a question mark because of the absence of any supporting matches to consensus splice sequences in the region of its internal S1 nuclease sensitive site. Transcript mapping of CLV and MSV did not reveal any evidence of spliced virus encoded RNAs (Morris-Krsinich *et al.*, 1985; Townsend *et al.*, 1985). However, introns of about the size mentioned above have been noted in plant genes, including those from the Gramineae (e.g. Dennis *et al.*, 1985; Werr *et al.*, 1985; Klosgen *et al.*, 1986; Paz-Ares *et al.*, 1987).

Transcripts 1− and 2− are not spliced in this region and contain the sequences which are absent in transcript 4−, suggesting that maturation of the virus specific nascent RNAs can be differentially regulated during the infection cycle. Different patterns of splicing of the same nascent RNA have been frequently reported for both the adenoviruses and papovaviruses (DePamphilis and Wasserman, 1982; Broker, 1984; Flint, 1986). Mature spliced and unspliced RNAs arising from the same transcription unit occur among the late mRNAs of the papovaviruses, the mRNAs of herpes simplex virus type 1 and from the integrated form of animal retroviruses (Broker, 1984). We are not aware of any reported cases of examples of selective processing of the transcripts of a single gene in plants, although such cases are documented for genes from mammals and *Drosophila* (Leff *et al.*, 1986).

The regulation of the splicing of the early mRNAs of the papovaviruses allows the synthesis, from the same transcription unit, of several proteins (the small, middle and large T antigens), which can serve different functions but share common sequences (DePamphilis and Wasserman, 1982). An analogous situation could exist for the expression of the complementary sense transcripts of DSV. Transcripts 2− (as well as 1− and 3−) could direct the synthesis of a 30.5 kd polypeptide and the spliced transcript 4− could code for a 41 kd polypeptide. Both polypeptides would share common N-terminal sequences and diverge at the junction between the 16.9 and 30.5 kd ORFs (Figure 4). The fusion of the 30.5 and 16.9 kd ORFs which is created in RNA 4−, retains all the conserved amino acid sequences identified upon comparison with the derived amino acid sequence of the ~40 kd ORF of the independently replicating component 1 (or A) of the whitefly transmitted sub-group of geminiviruses (Donson *et al.*, 1987). These complementary sense ORFs code for potential gene products which are the most conserved, at the amino acid level, throughout the whole geminivirus group, indicating that their function(s) may be essential for infection (Mullineaux *et al.*, 1985; Townsend *et al.*, 1985; Rogers *et al.*, 1986; Stanley *et al.*, 1986; Donson *et al.*, 1987). Like the T antigens of the papovaviruses, it is interesting to speculate that the products of the complementary sense transcripts of DSV could display a diversity of roles.

The most stable stem−loop structure on the DSV genome (hairpin 7; $\Delta G = -48$ kcal/mol; Donson *et al.*, 1987) and its nonanucleotide sequence 5′-TAATATTAC-3′ (as virion sense sequence) are transcribed as RNA 1− (Figure 6). The nonanucleotide sequence is conserved in the head of the most stable stem−loop structure in all geminiviruses so far sequenced (Davies *et al.*, 1987) and both features have been postulated to be involved in viral DNA replication, the regulation of transcription of the geminivirus genome and in packaging of virus ssDNA (Stanley, 1985; Davies *et al.*, 1987). The presence of such stable secondary structures in the leader sequence of RNA 1− could be inhibitory to the translation of the downstream 30.5 kd ORF, as demonstrated for a pre-proinsulin mRNA by Kozak (1986). This point plus the absence of any further amino acid coding capacity on RNA 1−, compared to that on RNA 2−, might suggest that it has a role other than translation. RNA 3−, like 2−, only has the potential to code for the 30.5 kd ORF. The 3′ end of RNA 3− is located very near to the 5′ end of the mapped position of the RNA moiety of the 'primer' found associated with encapsidated virus DNA (Donson *et al.*, 1987). Whether this observation is coincidental or of more significance remains to be established.

Based primarily on nucleotide sequence data, analogies have been drawn to the similarities in genome organization between the geminiviruses and the papovaviruses (Stanley and Gay, 1983; Howarth *et al.*, 1985; Townsend *et al.*, 1985; Davies *et al.*, 1987). In this study we have extended the analogy further by suggesting that a geminivirus can produce more than one RNA species (and presumably polypeptide) from the same transcription unit.

Materials and methods

Materials

All restriction endonucleases and DNA modifying enzymes were purchased from Gibco-BRL with the exception of the S1 nuclease which was bought from Sigma Chemical Co. Radioisotope labelled nucleotides were purchased from New England Nuclear.

Plants and virus

Digitaria setigera L. (described as *D.sanguinalis* by Dollet *et al.*, 1986) was obtained from Vanuatu. Plants were grown under the same conditions used for *Zea mays* L. by Mullineaux *et al.* (1988). DSV was inoculated into *Digitaria* plants, using a bacterial plasmid, pDS2 (Donson *et al.*, 1988) and the 'agro-infection' technique developed by Grimsley *et al.* (1987).

Molecular cloning

The recombinant M13 bacteriophages and the plasmid used for this work were constructed using standard recombinant DNA techniques (Maniatis *et al.*, 1982). The source of the DSV DNA used from the constructs was pDS1, which was known to code for a complete virus genome (Donson *et al.*, 1988).

RNA preparation

The RNA was prepared from healthy and infected *Digitaria* plants essentially as described by Morris-Krsinich *et al.* (1985) but with the modifications of Woolston *et al.* (1988). Preparation of poly(A)$^+$ RNA was as described by Maniatis *et al.* (1982).

Northern hybridizations and preparation of probes

These methods have been described previously (Morris-Krsinich *et al.*, 1985). High specific activity ^{32}P-labelled DNA probes were made using the Multiprime DNA labelling system from Amersham International, which is based on the method of Feinberg and Vogelstein (1984).

S1 nuclease mapping

Poly(A)$^+$ RNAs were mapped using the procedure of Berk and Sharp (1977) as described by Favaloro *et al.* (1980). The details of the conditions

used for the S1 nuclease mapping of geminivirus specific RNAs have been described previously (Morris-Krsinich *et al.*, 1985). The probe used to detect S1 nuclease resistant DNA fragments was the genome length *Hin*dIII DNA insert recovered from the dsDNA form of an M13 clone unless stated otherwise in the text.

High resolution mapping of the 5′ termini of transcripts 1+, 2+, 1−, 2−, 3−, 4− and 5− was carried out as described previously (Morris-Krsinich *et al.*, 1985), using appropriate terminal ^{32}P-labelled DNA restriction fragments prepared from pDSV1, a full length *Pst*I clone in pUC18. The protecting DNA fragments were prepared as follows: pDSV1 was digested with the first restriction endonuclease, end-labelled using T4 polynucleotide kinase (Maniatis *et al.*, 1982) and digested with a second restriction endonuclease. Following this pattern, those used were: (i) a 180 bp *Asp*718−*Hin*fI fragment (co-ordinates 253−73) for transcript 2+. (ii) A 244 bp *Hin*fI−*Nco*I fragment (co-ordinates 75−2532) for transcript 1+. (iii) A 217 bp *Sau*3A−*Nco*I fragment (co-ordinates 2318−2535) for transcripts 2− to 5− and (iv) a 323 bp *Nco*I−*Bgl*II fragment (co-ordinates 2532−154) for transcript 1−. Control treatments are described in the legend to Figure 4. The size of the S1 nuclease protected DNA fragments was determined using a sequence ladder from mp18 (Yanisch-Perron *et al.*, 1985) which was generated by the dideoxy chain termination procedure using methods previously cited in Mullineaux *et al.* (1984). For the sequencing reactions, the universal primer of Duckworth *et al.* (1981) was used, either labelled at the 5′ terminus with [γ-^{32}P]ATP (Maniatis *et al.*, 1982) and used with cold deoxy/dideoxy nucleotide mixes, or unlabelled but used in conjunction with [α-^{35}S]dATP in the nucleotide mixes. The sizes of the fragments were calculated using the corrections described by Donson *et al.* (1984).

The identification of DNA fragments corresponding to internal S1 nuclease sensitive sites in viral transcripts using two-dimensional agarose gel electrophoresis was according to the method of Favaloro *et al.* (1980). To detect such complementary sense RNAs, 5 μg poly(A)$^+$ RNA was used per protection experiment.

Acknowledgements

We would like to thank Dr J.W.Davies and all the Department of Virus Research for their interest and advice during the course of the work. We also thank Drs C.J.Woolston and J.Stanley for critical reading of the manuscript. We thank Ms Tarn Dalzell for secretarial assistance. This work was supported by the Agricultural and Food Research Council via a grant-in-aid to the John Innes Institute. G.P.A. and P.M.M. gratefully acknowledge the support of a NATO fellowship and a short-term EMBO fellowship respectively. This work was carried out under MAFF licence numbers 49/152 and 49A/41.

References

Abouzid,A.M., Frischmuth,T. and Jeske,H. (1988) *Mol. Gen. Genet.*, **212**, 252−258.

Andersen,M.T., Richardson,K.A., Harbison,S.-A. and Morris,B.A.M. (1988) *Virology*, **164**, 443−449.

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **12**, 721−732.

Broker,T.R. (1984) In Apirion,D. (ed.), *Processing of RNA*. CRC Press, Boca Raton, FL, USA, pp. 182−203.

Covey,S.N., Lomonossoff,G.P. and Hull,R. (1981) *Nucleic Acids Res.*, **9**, 6735−6747.

Davies,J.W., Stanley,J., Donson,J., Mullineaux,P.M. and Boulton,M.I. (1987) *J. Cell Sci., Suppl.*, **7**, 95−107.

Dennis,E.S., Sachs,M.M., Gerlach,W.L., Finnegan,E.J. and Peacock,W.J. (1985) *Nucleic Acids Res.*, **13**, 727−743.

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.

Dollett,M., Accotto,G.P., Lisa,V., Menissier,J. and Boccardo,G. (1986) *J. Gen. Virol.*, **67**, 933−937.

Donson,J., Morris-Krsinich,B.A.M., Mullineaux,P.M., Boulton,M.I. and Davies,J.W. (1984) *EMBO J.*, **3**, 3069−3073.

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.

Duckwork,M.L., Gait,M.J., Goelet,P., Hong,C.F., Singh,M. and Titmas,R.L. (1981) *Nucleic Acids Res.*, **9**, 1691−1706.

Favaloro,J.M., Treisman,R.H. and Kamen,R.I. (1980) *Methods Enzymol.*, **65**, 718−749.

Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **132**, 6−13.

Flint,S.J. (1986) *Adv. Virus Res.*, **31**, 169−228.

Grimsley,N., Hohn,T., Davies,J.W. and Hohn,B. (1987) *Nature*, **325**, 177−179.

Guilley,H., Dudley,R.K., Jonard,G., Balazs,E. and Richards,K.E. (1982) *Cell*, **30**, 763−773.

Hamilton,W.D.O., Stein,V.E., Coutts,R.H.A. and Buck,K.W. (1984) *EMBO J.*, **3**, 2197−2205.

Howarth,A.H., Caton,J., Bossert,M. and Goodman,R.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3572−3576.

Howell,S.H. (1984) *Nucleic Acids Res.*, **12**, 7359−7375.

Kozak,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2850−2854.

Klosgen,R.B., Gierl,A., Schwarz-Sommer,Z. and Saedler,H. (1986) *Mol. Gen. Genet.*, **203**, 237−244.

Lazarowitz,S.G. (1988) *Nucleic Acids Res.*, **16**, 230−249.

Leff,S.E., Rosenfeld,M.G. and Evans,R.M. (1986) *Annu. Rev. Biochem.*, **55**, 1091−1117.

MacDowell,S.W., Macdonald,H., Hamilton,W.D.O., Coutts,R.H.A. and Buck,K.W. (1985) *EMBO J.*, **5**, 1761−1767.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Messing,J., Geraghty,D., Heidecker,G., Hu,N.-T., Kridl,J. and Rubenstein, I. (1983) In Kasuge,T., Meredith,C.P. and Hollaender,A. (eds), *Genetic Engineering of Plants*. Plenum Press, New York, pp. 211−227.

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.

Mullineaux,P.M., Donson,J., Morris-Krsinich,B.A.M., Boulton,M.I. and Davies,J.W. (1984) *EMBO J.*, **3**, 3063−3068.

Mullineaux,P.M., Donson,J., Stanley,J., Boulton,M.I., Morris-Krsinich, B.A.M., Markham,P.G. and Davies,J.W. (1985) *Plant Mol. Biol.*, **5**, 125−131.

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.

Paz-Ares,J., Debabrota,G., Wienand,U., Peterson,P.A. and Saedler,H. (1987) *EMBO J.*, **6**, 3553−3558.

Rogers,S.G., Bisaro,D.M., Horsch,R.B., Fraley,R.T., Hoffman,N.L., Brand,L., Elmer,J.S. and Lloyd,A.M. (1986) *Cell*, **45**, 593−600.

Sharp,P.A., Konarska,M.M., Grabowski,P.J., Lamond,A.I., Marciniak,R. and Seiler,S.R. (1987) *Cold Spring Harbor Symp. Quant. Biol., Vol. LII*, 277−285.

Stanley,J. (1985) *Adv. Virus Res.*, **30**, 139−177.

Stanley,J. and Gay,M.R. (1983) *Nature*, **301**, 260−262.

Stanley,J., Markham,P.G., Callis,R.J. and Pinner,M.S. (1986) *EMBO J.*, **5**, 1761−1767.

Townsend,R., Stanley,J., Curson,S.J. and Short,M.N. (1985) *EMBO J.*, **4**, 33−38.

Werr,W., Frommer,W.-B., Maas,C. and Starlinger,P. (1985) *EMBO J.*, **4**, 1373−1380.

Woolston,C.J., Barker,R., Gunn,H., Boulton,M.I. and Mullineaux,P.M. (1988) *Plant Mol. Biol.*, **11**, 35−48.

Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103−119.

Received on November 14, 1988; revised on January 20, 1989