

# Transcript mapping reveals different expression strategies for the bicistronic RNAs of the geminivirus wheat dwarf virus

Elise L. Dekker, Crispin J. Woolston<sup>1</sup>, Yongbiao Xue<sup>2</sup>, Brian Cox<sup>2</sup> and Philip M. Mullineaux\*

John Innes Institute, John Innes Centre for Plant Science Research, Colney Lane,

Norwich NR4 7UH, <sup>1</sup>Department of Applied Biology, University of Hull, Hull HU6 7RX and

<sup>2</sup>Department of Plant Sciences, University of Oxford, South Parks Rd, Oxford OX1 3RA, UK

Received June 7, 1991; Revised and Accepted July 12, 1991

## ABSTRACT

**We have characterised the major transcripts of the Czech isolate of wheat dwarf virus (WDV-CJI) which show that WDV uses two different mechanisms for expressing overlapping open reading frames (ORFs). Mapping of the virion sense RNAs identified a single polyadenylated transcript of 1.1kb spanning the overlapping ORFs V1 and V2 which encode cell-cell spread functions and the coat protein respectively. This finding distinguishes WDV from other monocot-infecting geminiviruses studied so far which were shown to encode two 3' co-terminal transcripts capable of expressing either the V1 or V2 ORF. A survey of codon usage at the junction between the V1 and V2 ORF has led us to propose that translational frame shifting analogous to that in the yeast Ty element may occur. Analysis of polymerase chain reaction (PCR) amplified complementary sense cDNA clones has revealed the presence of mature spliced and unspliced RNAs which could encode products of an intron mediated C1:C2 ORF fusion or the C1 ORF product alone. Mapping of the 5' and 3' extremities of the major WDV encoded transcripts has allowed us to identify putative transcription regulatory sequences and the presence of multiple overlapping transcripts may suggest temporal regulation of transcription.**

## INTRODUCTION

The genomes of most plant viruses studied to date usually contain monocistronic segments which are transcribed as individual messages. Alternatively, all the genetic information contained within a given coding region is translated to give a single polypeptide which then is cleaved into individual proteins (1). An exception are the caulimoviruses which are thought to express polycistronic mRNAs by a relay-race model of translation (2).

The geminiviruses are a group of plant DNA viruses which have been proposed to be good models for studying aspects of

transcription and DNA replication (3). Geminivirus particles accumulate in the nuclei of infected cells where DNA replication and virus assembly probably take place (3). Their putative replicative forms are double-stranded (ds) covalently closed circular DNA of 2.6–3.0kb in chromatin-like structures (3,4). The mapping of geminivirus-specific polyadenylated transcripts for African cassava mosaic virus (ACMV; 5), maize streak virus (MSV; 6), tomato golden mosaic virus (TGMV; 7) and *Digitaria* streak virus (DSV; 8) revealed that their genomes are transcribed bidirectionally.

WDV possesses four open reading frames (ORFs; see Fig. 6), potentially encoding for polypeptides greater than 10kDa, which are common to the other monocot-infecting geminiviruses sequenced to date (9,10). In DSV and MSV, the V1 and V2 ORFs are spanned by the most abundant virus encoded RNAs detectable in infected plant material (6,8). In contrast, the C1 and C2 ORFs are encoded on mRNAs which are present at very low levels in infected material (6,8). DSV and a Swedish isolate of WDV have both been shown to splice their complementary sense RNA and it is likely that only the monocot-infecting geminiviruses display this form of RNA processing (8,11). In both cases, a functional intron has been located at the junction between C1 and C2 and its excision from a nascent RNA connects these two ORFs leading to the synthesis of a 41kDa C1:C2 fusion product (8,11). In addition, more abundant unspliced complementary sense RNAs have been demonstrated in DSV which would lead to the synthesis of a C1 encoded polypeptide of 30.5kDa (8).

Replication studies of an 'intronless' mutant of WDV in protoplasts of *Triticum monococcum* have shown that the C1:C2 fusion product encodes a protein which is required for WDV ds DNA replication (11). The role of the C1 product is unknown but is speculated to be involved in host specificity (11). The product of V1 of MSV has been shown to be involved in cell-cell spread of the virus and has been detected as a 10kDa polypeptide in protein extracts from infected leaves (12–14). The capsid protein is encoded by the V2 ORF (6) but is also required for systemic movement and leafhopper transmission (12,15).

\* To whom correspondence should be addressed

In this paper, we present the results of our studies on defining the major transcripts of WDV-CJI. This has revealed that WDV encodes two polypeptides for each transcription unit. We have unequivocally demonstrated the presence of spliced and unspliced mature RNAs which would lead to two different complementary sense polypeptides. We have been able to detect only one virion sense transcript spanning both V1 and V2 and we propose that a mechanism of translation frame shifting allows the synthesis of the V1 and V2 products from a single mRNA.

In addition to being a model for studying gene expression in the Gramineae, WDV has been used as a marker for gene transfer to cereal cells (16–18). Therefore, a complete transcript map of WDV will be of aid in the design of viral based vectors.

## METHODS

### Reagents

All restriction enzymes and DNA modifying enzymes were purchased from GIBCO BRL, New England Biolabs or Boehringer Mannheim, except Amplitaq (Taq polymerase) and Sequenase which were obtained from Perkin-Elmer Cetus and from US Biochemical Corporation respectively. Radiochemicals were obtained from New England Nuclear.

### RNA preparation and Northern blotting

WDV was agroinoculated into wheat plants (cv. Norman) as pJIT33 (15). Infected material for RNA was harvested 21 days post-inoculation. The preparation of total and poly (A)+ RNA from infected wheat tissue and its analysis by Northern blotting was carried out as previously described (8).

### Nuclease S1 mapping

The virion sense transcripts were mapped onto the WDV genome using the medium resolution nuclease S1 mapping procedures of Favaloro et al (19) as described previously (8). Genome length WDV DNA cloned in M13mp19 (20) at either the unique HindIII or KpnI sites (coordinates 2391 and 793 respectively; 9) was used as protecting ssDNA in the mapping experiments. The HindIII clone was anticipated to generate nuclease S1 protected fragments equivalent to the size of any virion sense transcript. The KpnI clone was expected to generate smaller fragments since the cloning site used is in the V2 ORF and is interrupted by M13 DNA.

The 5' terminus of the virion sense transcript was determined by a modified nuclease S1 mapping procedure (21). The protecting 'probe' DNA was a WDV HpaII fragment (coordinates 2611–318) in which the overhang had been filled-in with T4 DNA polymerase I and then cloned into the SmaI site of M13mp18 (20). The nature of this construct was confirmed by sequencing. Labelled, single-stranded DNA was synthesised from this template and hybridised with 1 µg of poly (A)+ RNA from WDV infected leaf tissue. Subsequent nuclease S1 digestion yielded a protected, labelled DNA fragment with its 5' and 3' ends corresponding to the 5' end of the transcript and the WDV:M13 cloning junction respectively. This fragment was sized on a 6% (w/v) polyacrylamide sequencing gel against a di-deoxy sequence ladder of a SmaI-SstI clone which had been digested with HpaII prior to electrophoresis. Due to the in-filling of the HpaII site in the protecting DNA (+3bp) and the lack of a 5' hydroxyl group on nuclease S1 treated fragments (–0.5bp), the position of the protected fragment relative to the sequence ladder must be adjusted by 2.5bp.

The 3' extremity of the virion sense transcript was mapped using the procedure of Berk and Sharp (22). The <sup>32</sup>P-labelled protecting DNA fragment was prepared by cutting 10 µg of pJIT43 (9) with MluI (coordinate 1246; 9) end-labelled with <sup>32</sup>P dC-TP and Klenow fragment of DNA polymerase I as previously described (8). The end-labelled DNA was then cut with RsaI (coordinate 1431; 9), subjected to electrophoresis through a 5% (w/v) polyacrylamide gel and the 185bp band located by autoradiography and eluted (23). Nuclease S1 protected fragments were separated as for the 5' mapping procedure (21).

### Primer extension

Preparations of total RNA (100 µg) extracted from WDV-infected plants, healthy plants and RNA treated with DNAase-free RNAase A were primed with 10 pmols of a <sup>32</sup>P 5' end-labelled synthetic oligomer (5'-GATGAAGAGGCCATGGTAGTGAA-CAGAAGTCCGGC3', coordinates 2581–2615 [9]; specific activity of 1.2 × 10<sup>5</sup> cpm/pmol) and extended along the RNA template with MoLV reverse transcriptase as described by Sambrook et al (24). Primer extension products were resolved on an 8% (w/v) polyacrylamide-urea gel (8) and sized by comparison with a dideoxysequencing ladder primed with the same oligomer onto WDV sequences in pJIT43 (10) using the plasmid sequencing method of Zhang et al (25).

### PCR amplification

Analysis of rare WDV-specific complementary sense transcripts was carried out using the polymerase chain reaction (PCR) 3'RACE (rapid amplification of cDNA ends; 26). The use of this procedure to map geminivirus complementary sense transcripts has been described (27) and is shown in Fig. 3. The PCR reaction conditions and cycles, analysis of the amplified cDNA by cutting with restriction enzymes and Southern blotting have been described previously (27). The WDV-specific primer employed was from coordinates 2151–2176 (5'-CAGATCATTGAATC-TAGTTCCTCTCGC-3'; primer #1 in Fig. 3). The fidelity of the primer was checked by its ability to prime the expected sequences in pJIT43 using a plasmid sequencing procedure (25). Unless otherwise stated a genome length HindIII fragment from pWDVK10D (15) was used as the WDV specific probe.

### Molecular cloning and sequencing of 3'RACE products

PCR amplified cDNA was digested with EcoRI (coordinate 2142; Fig. 3) and SalI, for which a restriction site was present in the dT<sub>17</sub>-adaptor sequence (26; Fig. 3). The restricted DNA was electro-eluted from agarose gels and cloned into the corresponding restriction sites of M13mp19 (20). WDV positive plaques were identified by plaque hybridisation using a WDV-specific probe corresponding to coordinates 1246–2242. Dideoxysequencing was carried out as described previously (9) using the universal primer of Duckworth et al (28) to sequence across the putative intron-exon splice junction previously described for the Swedish isolate of WDV (WDV-S; 11). A synthetic primer (5'-TGGTT-TGTCTTTGCTCGCTAGCCG-3'; coordinates 1326–1348) was used to sequence across the 3' ends of the cDNA clones. Several clones could not be sequenced employing this specific primer because their 3' ends were upstream of coordinate 1326 (see Results). These clones were inverted by cloning their EcoRI-SalI fragments from M13mp19 into M13mp18 (20), followed by dideoxysequencing across the SalI cloning junction using the universal primer.

## RESULTS

### Virion sense transcript mapping

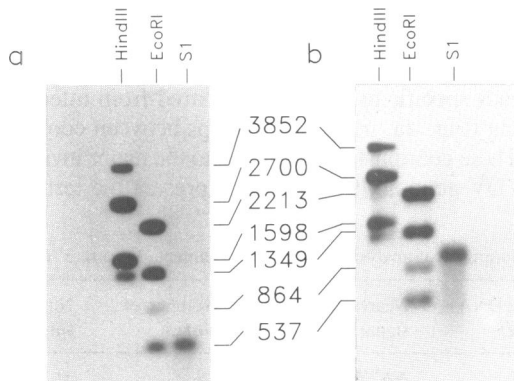
Northern blot analysis of poly (A)+ RNA using a WDV virion strand specific probe revealed the presence of one RNA species of ca. 1.1kb specific to infected tissue (data not shown). Medium resolution nuclease S1 mapping (see Methods) using protecting complementary sense ssDNA of WDV cloned outside the virion sense transcription unit revealed one protected fragment of 1060 bases (Fig. 1b). Using ss WDV M13 DNA cloned at the KpnI site (coordinate 793: 9) in the V2 ORF resulted in two co-migrating protected DNA fragments of ca. 530 bases (Fig. 1a). Therefore, the 5' and 3' ends of the virion sense RNA were mapped approximately to coordinates 240 and 1280 respectively. No nuclease S1 protected fragment could be detected when poly (A)+ RNA from healthy tissue was used (data not shown).

High resolution nuclease S1 mapping revealed the major 5' end of the virion sense RNA to be at coordinate 247 (band 1; Fig. 2b) and two lesser abundant protected fragments mapping to coordinates 243 and 248 (bands 2a and b respectively; Fig. 2b). It is unusual for the major start site of a transcript to start with a C. It remains possible that a GC rich region immediately 5' to the RNA end, influenced the relative abundance of protected fragments. Using the same technique (see Methods), a single S1 nuclease protected band corresponded to the 3' end of the virion sense transcript mapped to coordinate 1302 (data not shown).

### Complementary sense transcript mapping using PCR

Because of their low abundance, attempts at analysing complementary sense transcripts of WDV by nuclease S1 mapping (8) were unsuccessful. Therefore, we employed the 3'RACE PCR procedure (see Methods) to amplify complementary sense specific cDNA and map the 3' ends of spliced and unspliced RNAs (Fig. 3).

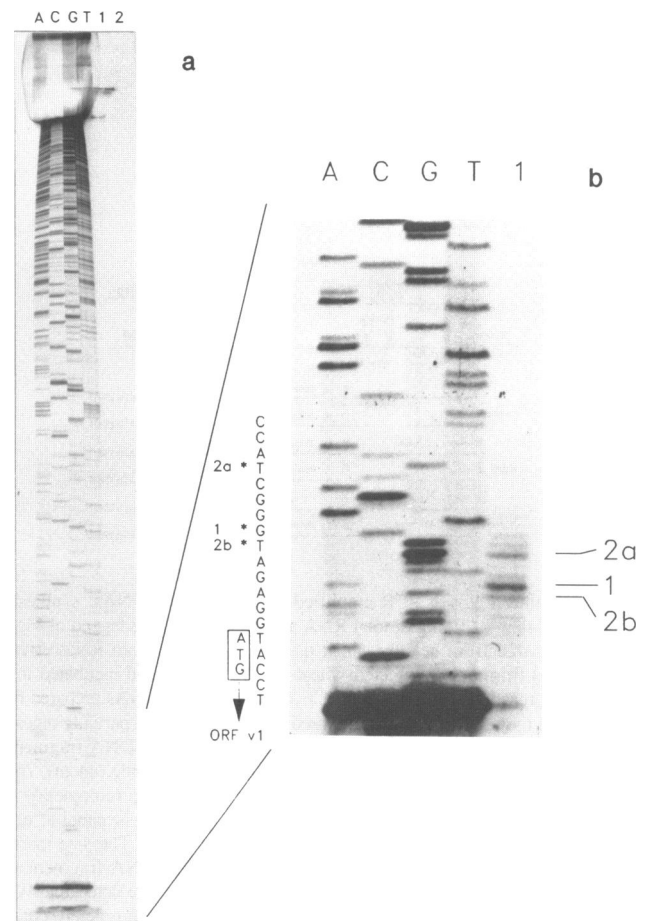
Southern blot analysis of the uncut cDNA obtained after PCR amplification using primer #1 (Fig. 3 and Methods) and the adaptor primer (Fig. 3 and Methods) revealed a major band of ca. 0.9kb (band a) and a minor band of ca. 0.8kb (band b; Fig. 3a track 3). No WDV specific bands were detected in samples of



**Figure 1.** Southern blot of nuclease S1 resistant fragments of complementary sense M13 ssDNA clones (which gives a virion sense sequence ladder) protected by WDV virion sense RNA (tracks labelled S1). The M13 clones used were full length WDV DNAs cloned at either the single KpnI site in the V2 ORF (a) or the single HindIII site in the C1 ORF (b; see Methods). The size ladder is given in base pairs between the panels and was generated by digesting pWDVK10D (15) with the enzymes indicated above each track. The probe was a full length, oligo-primed <sup>32</sup>P-labelled WDV-CJI KpnI DNA fragment (see Methods).

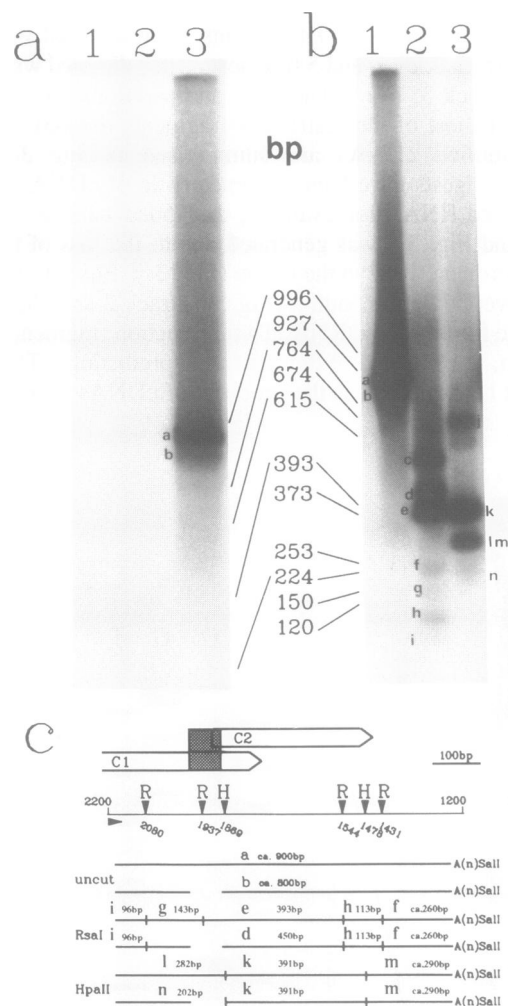
PCR amplified cDNAs prepared from either RNA of healthy leaf tissue (Fig. 3a; track 1) or from RNAase treated RNA from infected material (Fig. 3a; track 2).

The PCR amplified cDNA clones were mapped onto the WDV genome using the restriction enzymes RsaI and HpaII (Fig. 3b and c). The WDV specific cDNAs (bands a and b; Fig. 3a, track 3; Fig. 3b, track 1) gave 7 fragments when digested with RsaI (Fig. 3b, track 2; c-i) and 5 fragments when digested with HpaII (Fig. 3b, track 3; j-n). The sizes and positions on the WDV genome of most of the restriction fragments derived from the PCR amplified cDNAs are summarised in Fig. 3c. Both restriction digests gave bands characteristic of cDNAs derived from spliced RNA. For example, the 450bp band d (Fig. 3b, track 2 and Fig. 3c) was generated due to the loss of the RsaI site at coordinate 1937 in the intron (Fig. 3c). Equally, the same digest gave the 393bp band 3 (Fig. 3b, track 2 and Fig. 3c) as diagnostic for unspliced RNA. Two restriction fragments (c and j; Fig. 3b, tracks 2 and 3) did not fit the predictions. The bands could not be explained as the products of cDNAs derived from



**Figure 2.** Location of the 5' end of the WDV virion sense transcript using the method of Aldea *et al* (21; see Methods). Track 1 (a and b) contains the nuclease S1 resistant fragments (bands 1, 2a and 2b) generated from mixing the protecting DNA fragment (HpaII; see Methods) with 1µg of poly (A)+ RNA from WDV infected wheat leaves. Track 2 (a) is the nuclease S1 treated protecting DNA without the addition of RNA. The size ladder is the sequence of a WDV SmaI-SstI clone digested with HpaII, to correspond to the protected DNA. The position of the bands on the annotated sequence has been adjusted by 2.5bp (see Methods). The products were separated on a 6% (w/v) polyacrylamide-urea sequencing gel as previously described (8). The gel was dried and subjected to autoradiography.

RNA which had longer or shorter 3' ends than those depicted in Fig. 3c. Nor were they the products of primer # 1 binding at alternative 5' sites, since these would have been present as cDNA larger than band a in Fig. 3a (track 3) and Fig. 3b (track 1). Since



**Figure 3.** Southern blot of PCR amplified WDV specific cDNAs, synthesised from the equivalent of 100ng of total RNA and probed as in the legend of Fig. 1. **a.** cDNA prepared from healthy wheat tissue (track 1) or from RNase treated RNA prepared from WDV infected wheat tissue (track 2) and incubated in the 3'RACE PCR reactions. Track 3. 3'RACE PCR amplified cDNAs prepared from WDV infected tissue. The size markers used were generated by partially digesting the 996bp (WDV coordinates 1246–2242; 9,10) MluI-PvuII DNA fragment of pJIT43 (10) with HpaII, generating WDV specific bands of 996bp (uncut), 764bp (partially cut), 615bp (partially cut), 393bp, 373bp and 224bp (completely cut) in size. **b.** WDV specific 3'RACE PCR amplified cDNA as in panel a, either uncut (track 1) or cut with RsaI (track 2) or HpaII (track 3). The size markers used were generated by partially digesting the MluI-PvuII DNA fragment with HpaII giving the bands as in panel a and with Sau3A to generate detectable WDV specific bands of 927bp (partially cut), 674bp and 253bp (completely cut). **c.** The WDV genome between coordinates 2200 and 1200 (9,10). The shaded area indicates the position of the intron in the WDV-S genome (11) and the positions of the C1 and C2 ORFs are shown (see also Fig. 5). Below the ORFs, the coordinates of the restriction enzymes RsaI (R) and HpaII (H) are indicated. The priming site and the direction of second strand synthesis of oligonucleotide # 1 (see Methods) used to generate the PCR amplified cDNAs is marked (▶). Below the coordinates are the sizes of the restriction fragments d-n (excluding j) shown in panel b and the interpretation of how they are generated by digesting either spliced or unspliced PCR-amplified cDNA. The unspliced and spliced cDNAs are depicted as the top and bottom of each pair respectively. For an explanation of bands c and j in panel b see Results. A(n)Sall is the adaptor primer also used in the PCR reaction in conjunction with oligonucleotide # 1 (see Methods).

no examples of altered restriction patterns could be found among the cloned cDNAs (see below) no further characterisation of these bands was attempted. A similar unexplained restriction fragment generated from PCR amplified complementary sense cDNA derived from DSV encoded RNA has been described which was thought to be associated with amplification of the spliced transcript (27).

### Molecular cloning and sequence analysis of PCR amplified cDNAs

PCR amplified complementary sense WDV cDNA was cut with EcoRI (coordinate 2142) and Sall (in adaptor: Fig. 3c) and ligated into the EcoRI and Sall sites of M13mp19 (see Methods: 20).

The results of the sequence analysis of the positive clones recovered are shown in Table 1. As reported previously (27) only ca. 10% of clear (lac-) plaques contained WDV specific inserts. The sequencing of the 3' ends of most of the clones was carried out using an oligonucleotide primer between coordinates 1326–1348 (see Methods). Only cDNA clones containing a poly(A) tail were included in the analysis. Those without poly(A) tails accounted for a further 9% of the clones. Eight clones had 3' ends upstream of the specific primer (Table 1) and were identified after sub-cloning and sequencing in M13mp18 (see Methods). 22.6% of the cDNAs were derived from RNA spliced between coordinates 1964 and 1876 (Table 1). The 86bp WDV intron has been previously shown to have 5' donor and 3' acceptor sites which agree with consensus sequences (11).

### Mapping the 5' termini of the complementary sense transcripts

Primer extension was used to map the 5' termini of complementary sense transcripts, although the amounts of RNA and primer used were at the top end of the recommended range (see Methods; 24). The two most abundant primer extension products unique to RNA prepared from infected tissue (Fig. 4a; track 4) corresponded to coordinates 2690 and 2693 on the WDV-CJI genome (Fig. 4a and b; track 4). RNA prepared from healthy tissue also gave prominent primer extension products (Fig. 4a; track 3) and may represent host RNAs with similar sequences. Similar to their counterparts in WDV, the major complementary sense transcripts of DSV also have two apparent 5' ends separated by only a few nucleotides (8).

In addition to the two main primer extension products, a faint set of bands specific to the RNA prepared from infected tissue, is apparent (Fig. 4a, track 4) and maps between coordinates 30 and 40. These coordinates correspond to the major inverted repeat of the WDV genome (9) and may represent the termination of

**Table 1.** Summary of analysis of WDV complementary sense 3' cDNA clones.

3' end of cDNA (coordinate)	Nearest poly (A) signal	Number of clones	Number spliced
1277	AATAAA: 1293	9	3
1280	AATAAA: 1293	14	5
1347	TATAAT: 1354	5	0
1407	TATAAA: 1422	1	0
1968	N.D.	1	0
2002	N.D.	1	0

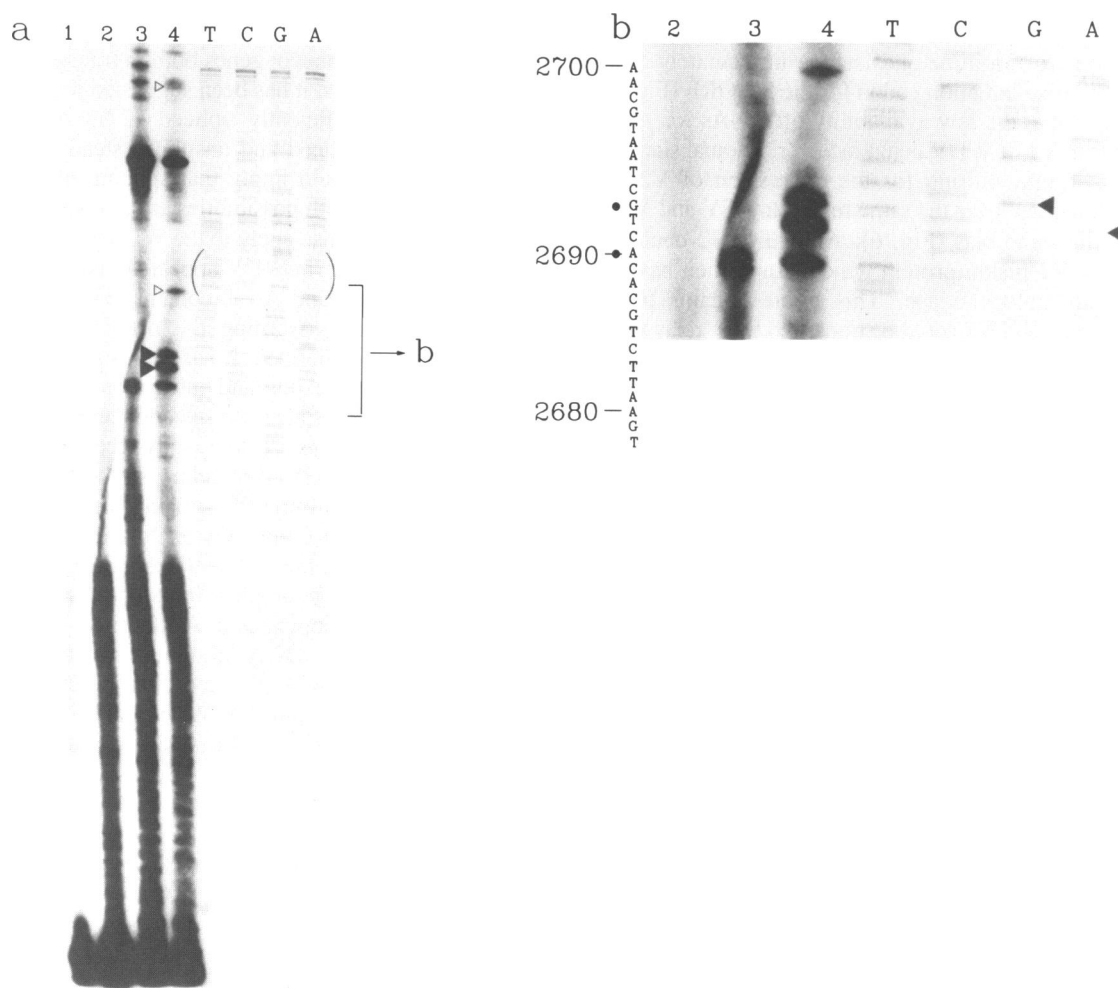
The coordinates are from virion sense strand of the WDV-CJI sequence (10). The consensus of the polyadenylation sequence of animal genes (AATAAA) is from Proudfoot and Brownlee (30). Putative alternative polyadenylation signals also shown are from Dean *et al* (44). N.D. = none detected.

the reverse transcriptase at a major stem-loop structure present in the RNA template. The position of the inverted repeat is also recognisable by the stacking of the bands of the dideoxy-sequencing ladder in that region (Fig. 4a; tracks T,C,G,A). These data indicate that a minor RNA species exists which extends to at least coordinate 40 and may be equivalent to the RNA 1- of DSV whose 5' end (determined by nuclease S1 protection) is 98bp beyond the 3' end of the conserved stem-loop (8).

## DISCUSSION

The major virus encoded transcripts whose extremities have been mapped onto the WDV-CJI genome are shown in Fig. 5. The 5' ends of the major transcripts encoded by both the virion and complementary sense strands map 17 bases and 30 bases respectively downstream of the consensus 'TATA' sequence

commonly found in eukaryotic promoters (TATAA/TAA/T; 29; Figs. 4 and 5). Equally, consensus polyadenylation sequences (AATAAA; 30) are located immediately upstream of the predominant 3' ends of the virion and complementary sense RNAs (Table 1 and Fig. 5). Thus, the WDV genome is transcribed bidirectionally and has an arrangement of transcription regulatory sequences in common with those of other geminiviruses whose transcripts have been mapped onto their genomes, especially those of the sub-group which infect the Gramineae (6-8). However, it should be noted that attempted alignment of the large intergenic regions of the Gramineae infecting geminiviruses results in no extensive blocks of sequence homology being identified, save for the head of the stem-loop structure (5'TAATATTAC3'; 3) and the TATA boxes (P.M. Mullineaux; unpublished data). This is despite this sub-group of geminiviruses often sharing common hosts and the functional

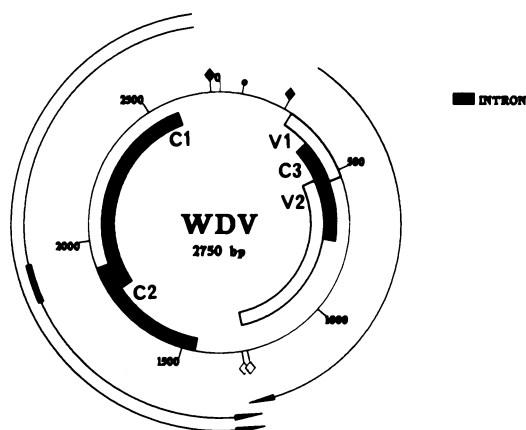


**Figure 4.** Location of the 5' ends of the major WDV encoded complementary sense RNAs by primer extension (see Methods). The primer used is located on the WDV sequence between coordinates 2581 and 2615. 10 pmols of primer ( $1.2 \times 10^5$  cpm/pmol) and (where appropriate) 100  $\mu$ g of total RNA were used per reaction. The products of the primer extension reactions (25,000 cpm/track) were separated on an 8% (w/v) sequencing gel and visualised as described in the legend of Fig. 2. The tracks are as follows: Primer only (track 1), primer extension reactions carried out with RNAse treated RNA from WDV infected tissue (track 2), with RNA from healthy tissue (track 3) and from infected tissue (track 4). The same primer was used to generate the sequencing ladder (TCGA) from pJIT43 (see Methods; 10). The two most prominent bands are marked in panel a ( $\blacktriangleright$ ) and are downstream from a consensus eukaryotic TATA box (29) marked in parentheses on the sequencing ladder (5'-TATATAA-3'; coordinates 2719-2713). The less prominent extension products discussed in the text are marked ( $\triangleright$ ). The area bounded by the bracket to the right of panel a is magnified in panel b. The bands in the sequencing ladder which co-migrate with the major primer extension products in track 4 are marked ( $\circ$ ). It should be noted that the bands marked migrate equivalent to 0.5 nucleotides slower than their corresponding primer extension band because they contain a hydroxyl group rather than a phosphate group at their 5' ends. To the left the virion sense sequence is shown and the coordinates corresponding to the 5' end of the complementary sense RNA are marked ( $\blacktriangleleft$ ).

identification of certain cis motifs as Sp1-like transcription factor binding sites in the virion sense promoter of MSV (31,32).

Nuclease S1 mapping of the virion sense RNAs has revealed the presence of a single transcript which spans both the V1 and V2 ORFs (Figs. 1 and 5). This is in contrast to virion sense transcription in MSV and DSV (6,8), where two 3' co-terminal virion sense transcripts are found with a 5' end either upstream of V1 or V2. Thus in DSV and MSV, V1 and V2 are each present as the most 5' cistron in a separate RNA species. In contrast, the single virion sense RNA of WDV may be used to translate both V1 and V2. One possible mechanism which would allow both V1 and V2 to be expressed from one RNA would be translational frame shifting at the junction between the two ORFs, analogous to that occurring in the Ty retrotransposon of yeast (33, 34). It has been suggested that an imbalance between the cellular tRNA pool and the pattern of codon usage at the junction between two overlapping ORFs signals the occurrence of frame shifting (35). Examination of the codon usage at the junction of V1 and V2 in WDV, DSV and MSV revealed the presence of rarely used codons for arginine and leucine immediately 5' to the putative methionine initiation codon for the V2 ORF (Fig. 6; 9,10). We suggest that the low availability of tRNAs for AGA and UUA codons in V1 of WDV could induce ribosomal slippage and cause a +2 frame shifting for the expression of V2. In contrast, the codon usage over the same region in DSV and MSV did not display the same bias (Fig. 6). As in Ty, V2 could be produced as a V1:V2 fusion protein which would then have to be cleaved by an endopeptidase. This point remains to be established. Alternatively, V2 could be translated by a 'relay race' type of ribosome frameshifting as proposed for CaMV (2, 36).

In MSV, virion sense ORFs have been shown to be essential for systemic spread and expression of symptoms (12, 13) and V2 encodes the coat protein (6). In WDV, deletion of both the V1 and V2 ORFs does not affect the ability of the virus to replicate as ds DNA, but abolishes the ability of the virus to spread systemically (15, 18). The high degree of conservation of the derived amino acid sequences of the V1 and V2 ORFs emphasises that these sequences are expressed and probably have the same functions (9, 14). The reason for such a radical difference in the expression of V1 and V2 of MSV and DSV



**Figure 5.** The major virus specific transcripts mapped onto the WDV-CJI genome. The positions of the ORFs, most likely TATA sequences (♦), most likely polyadenylation sequences (◊) and the conserved, potentially most stable stem loop structure are from MacDowell *et al* (9) and Woolston *et al* (10). The position of the intron is marked by the black arc dividing the spliced transcript.

on the one hand and WDV on the other hand is not apparent.

Mature spliced and unspliced polyadenylated transcripts spanning the C1 and C2 ORFs have been mapped and their PCR amplified cDNAs cloned and sequenced (Figs. 3 and 5; Table 1). The 3'RACE procedure used allows the study of cDNAs derived from RNA and not contaminating geminivirus DNA and is particularly important in studying cDNAs containing introns (27). Thus, we have been able to confirm that like WDV-S (11), WDV-CJI possesses an 86bp intron between coordinates 1876 and 1964, using splice site junctions that match consensus donor and acceptor sequences (37). The removal of the intron results in a spliced RNA leading to the fusion of the C1 and C2 ORFs (Fig. 5) which would allow it to potentially direct the synthesis of an approximately 41kDa polypeptide.

The presence of mature spliced and unspliced complementary sense transcripts has now been recorded for both WDV and DSV (8, 27) and may be a feature of the complementary sense transcription unit of all monocot-infecting geminiviruses (11). From a comparison of the functioning of the DSV intron and introns of several monocot genes both in transgenic tobacco and in their native species, it has been suggested that the DSV intron is intrinsically inefficiently spliced (27). Alternatively, the presence of both spliced and unspliced steady state transcripts in infected tissue could imply that splicing of the DSV intron is controlled (27). Both possibilities are not necessarily mutually exclusive and could also apply to the WDV intron. Like the DSV intron, its counterpart in WDV appears to be typical of monocot introns in both size and A+T content (53%; 38–40).

The mechanisms governing the occurrence of differential splicing are not well defined. It has been suggested that both cis-acting sequences in exons and introns as well as trans-acting factors can influence splice site selection and splicing efficiency (38). If so, such features are not readily recognisable in the intron of WDV or that of DSV. Secondary structure may also play a role in unregulated differential splicing, involving pre-mRNAs adopting one of several mutually exclusive conformations (37).

To our knowledge, the differential splicing observed for both WDV and DSV is the simplest form of intron skipping so far described. Intron skipping is a widely used mechanism in the processing of nascent RNAs of animals and their DNA viruses (37, 41, 42) and allows multiple protein products to be derived from one transcription unit. For WDV and DSV this ensures that both the C1:C2 and C1 products can be made.

WDV	
14 23 38 9 3 25 18 73 54 54 17	
T P A R L N G D Q Q G	V1
ACACCGCCAGATTAAATGGTGACCAACAAGGA	
* M V T N K	V2
100 36 46 75 86	
DSV	
19 25 26 33 28 34 30 15 33 4	
T E S H L P C L E R *	V1
ACTGAAAAGCCATCTGCCATGCTTTCATCGATGA	
* K P S A M S S S M	V2
86 34 15 38 100 15 11 14 100	
MSV	
25 26 21 34 25 19 34 44 21 17	
H P G P F V P G T G *	V1
AATCCCGGCCATTTGTTCCAGGCACGGGATAA	
I P G H L F Q A R D	V2
65 23 44 33 14 75 46 16 13 27	

**Figure 6.** Codon usage frequency at the junction between the V1 and V2 ORFs of three geminiviruses which infect the gramineae. The numbers above and below the derived amino acid sequences encoded by the V1 and V2 ORFs respectively indicate the percentage of occurrence for each codon of a given amino acid according to the calculations by Murray *et al* (45).



In addition to the major complementary sense RNAs described, cDNAs have been cloned which may represent a minor population of RNAs with highly heterogeneous ends terminating up to 725 bases upstream of a major 3' end. A population of complementary sense RNAs with variable 3' ends has also been identified from DSV infected *Digitaria* but not from tobacco transgenic for DSV DNA (27). These RNA species may be associated only with replicating viral DNA and may represent a failure of the host transcription machinery to interact correctly with the replicating viral genome (27). This minor heterogeneity of the 3' ends of the complementary sense RNAs is consistent with the view that the polyadenylation sequences commonly used by plant genes can be highly flexible with putative polyadenylation signals up to 130 bases upstream from the 3' end of the transcript (43, 44).

The primer extension mapping also revealed the existence of a minor RNA species which extended at least as far as coordinate 40. This suggests the presence of a minor RNA species, perhaps equivalent to RNA 1- of DSV, as identified by nuclease S1 mapping (8). The function of such a transcript in either WDV or DSV remains obscure, but it has been previously noted that the presence of a stable stem-loop structure in the untranslated leader sequence of RNA 1- would abolish translation of the C1 ORF and that these RNA species may not be translated (8).

## ACKNOWLEDGEMENTS

This work was supported by a grant-in-aid from the AFRC to the John Innes Institute. E.L.D. and C.J.W. gratefully acknowledge the support of a long-term EMBO Fellowship and the DTI sponsored Plant Gene Tool Kit Consortium respectively. We thank Julie Hofer and Helen Reynolds for critical reading of the manuscript. This work was carried out under MAFF licence number PHF 1185/89 (89).

## REFERENCES

- Hull, R. (1990) *Seminars in Virol.*, **1**, 239–247.
- Seig, K. and Gronenborn, B. (1982) *NATO/FEBS Advanced course: Structure and function of plant genomes*. p. 154.
- Davies, J.W., Stanley, J., Donson, J., Mullineaux, P.M. and Boulton, M.I. (1987) *J. Cell Sci. Suppl.*, **7**, 95–107.
- Abouzid, A.M., Frischmuth, T. and Jeske, H. (1988) *Mol. Gen. Genet.* **212**, 252–258.
- Townsend, R., Stanley, J., Curson, S.J. and Short, M.N. (1985) *EMBO J.* **4**, 33–38.
- 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.
- Sunter, G., Gardiner, W.E. and Bisaro, D.M. (1989) *Virology*, **170**, 243–250.
- Accotto, G.P., Donson, J. and Mullineaux, P.M. (1989) *EMBO J.* **8**, 1033–1039.
- MacDowell, S.W., MacDonald, H., Hamilton, W.D.O., Coutts, R.H.A. and Buck, K.W. (1985) *EMBO J.* **4**, 2173–2180.
- Woolston, C.J., Barker, R., Gunn, H.V., Boulton, M.I. and Mullineaux, P.M. (1988) *Plant Mol. Biol.* **11**, 35–43.
- Schalk, H.-J., Matzeit, V., Schiller, B., Schell, J. and Gronenborn, B. (1989) *EMBO J.* **8**, 359–364.
- Boulton, M.I., Steinkeller, H., Donson, J., Markham, P.G., King, D.I. and Davies, J.W. (1989) *J. Gen. Virol.*, **70**, ?
- Lazarowitz, S.G., Pinder, A.J., Damsteegt, V.D. and Rogers, S.G. (1989) *EMBO J.* **8**, 1023–1032.
- 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.
- Woolston, C.J., Reynolds, H.V., Stacey, N.J. and Mullineaux, P.M. (1989) *Nucleic Acids Res.* **17**, 6029–6041.
- Topfer, R.J., Gronenborn, B., Schell, J. and Steinbiss, H.H. (1989) *Plant Cell* **1**, 133–139.
- Creissen, G., Smith, C., Francis, R., Reynolds, H. and Mullineaux, P. (1990) *Plant Cell Reports* **8**, 680–683.
- Ugaki, M., Ueda, T., Timmermans, M.C.P., Vieira, J., Elliston, K.O. and Messing, J. (1991) *Nucleic Acids Res.* **19**, 371–377.
- Favaloro, J.M., Treisman, R.H. and Kamen, R.I. (1980) *Methods Enzymol.* **65**, 718–749.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* **33**, 103–119.
- Aldea, M., Claverie-Martin, F., Diaz-Torres, M.R. and Kushner, S.R. (1988) *Gene* **65**, 101–110.
- Berk, A.J. and Sharp, P.A. (1977) *Cell* **12**, 721–732.
- Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edition. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Zhang, H., Scholl, R., Browse, J. and Somerville, C. (1988) *Nucleic Acids Res.* **16**, 1220.
- Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.
- Mullineaux, P.M., Guerineau, F. and Accotto, G.-P. (1990) *Nucleic Acids Res.* **18**, 7259–7265.
- Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmas, R.L. (1981) *Nucleic Acids Res.* **9**, 1691–1706.
- Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* **50**, 349–383.
- Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* **236**, 211–214.
- Fenoll, C., Black, D.M. and Howell, S.H. (1988) *EMBO J.* **7**, 1589–1596.
- Fenoll, C., Schwarz, J.J., Black, D.M., Schneider, M. and Howell, S.H. *Plant Mol. Biol.* **15**, 865–877.
- Belcourt, M.F. and Farabaugh, P.J. (1990) *Cell* **62**, 339–352.
- Xu, H. and Boeke, J.D. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8360–8364.
- Kozak, M. (1986) *Adv. Virus Res.* **31**, 229–292.
- Dixon, L.K. and Hohn, T. (1984) *EMBO J.* **3**, 2731–2736.
- Green, M.R. (1986) *Ann. Rev. Genet.* **20**, 671–708.
- Goodall, G.J. and Filipowicz, W. (1990) *Plant Mol. Biol.* **14**, 727–733.
- Goodall, G.J. and Filipowicz, W. (1989) *Cell* **58**, 473–483.
- Hanley, B.A. and Schuler, M.A. (1988) *Nucleic Acids Res.* **16**, 7159–7175.
- Krainer, A.R. and Maniatis, T. (1988) In: *Frontiers in Molecular Biology: Transcription and splicing*. Hames, B.D. and Glover, D.M. eds. pp. 131–206. IRL Press, Oxford and Washington D.C.
- Laski, F.A., Rio, D.C. and Rubin, G.M. (1986) *Cell* **44**, 7–19.
- Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1985) *Nucleic Acids Res.* **14**, 2229–2240.
- Joshi, C.P. (1987) *Nucleic Acids Res.* **15**, 9627–9640.
- Murray, E.E., Lotzer, J. and Eberle, M. (1989) *Nucleic Acids Res.* **17**, 477–498.