Journal of General Virology (1998), **79**, 3111–3118. Printed in Great Britain

Sequences of ten circular ssDNA components associated with the milk vetch dwarf virus genome

Yoshitaka Sano,[1] Masafumi Wada,[2] Yoshifumi Hashimoto,[1] Tsuguo Matsumoto[1] and Makoto Kojima[2]

[1] Department of Applied Biology, Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto 606-8585, Japan

[2] Lab. Plant Pathology, Faculty of Agriculture, Niigata University, Ikarashi 2-8050, Niigata 950-2102, Japan

Milk vetch dwarf virus (MDV) is a member of the proposed genus *Nanovirus*, and its genome is composed of multiple, circular ssDNA components of about 1 kb. We have cloned and sequenced ten ssDNA components and designated them MDV-C1 to C10. Each DNA component contains one potential major open reading frame, and contains a putative stem–loop structure in the non-coding region. Notably, four components (C1, C2, C3 and C10) encode distinct replication-associated (Rep) proteins of 33 kDa, which show only limited (42–57%) amino acid identity. The six other components encode proteins with calculated molecular masses ranging from 12·7 to 19·7 kDa. Comparison of the sequences with those of other nanoviruses reveals that MDV is closely related to faba bean necrotic yellows virus (FBNYV) and subterranean clover stunt virus (SCSV). Six putative MDV genome products, including one Rep and five non-Rep proteins, show high (70·4–90·9%) amino acid identity to the corresponding six FBNYV proteins, whereas two other Rep proteins encoded by MDV-C2 and C3 are 82·3% and 73·0% identical to those encoded by SCSV-C2 and C6, respectively. These results indicate that MDV, FBNYV and SCSV have diverged from a common origin, which had multiple Rep components. In addition, the putative proteins encoded by MDV-C4 and its homologues contain a consensus retinoblastoma-binding motif, suggesting that they may be involved in controlling the host cell cycle.

Introduction

Milk vetch dwarf virus (MDV) is an aphid-borne virus that infects several legume crops in Japan. The name is derived from Chinese milk vetch (*Astragalus sinicus* L.), a common green-manure crop, from which the disease was first reported (Matsuura, 1953; Inouye *et al.*, 1968). The virus is transmitted by *Aphis craccivora* in a persistent manner, and causes yellowing and dwarfing of Chinese milk vetch, broad bean (*Vicia faba* L.), pea (*Pisum sativum* L.) and soybean [*Glycine max* (L.) Merr.] plants. MDV was originally thought to be a luteovirus because of its symptoms, vector transmission and the observation of 26 nm isometric particles in MDV-infected broad bean leaves (Ohki *et al.*, 1975). However, isometric particles measuring 18 nm in diameter and containing ssDNA of about 1 kb were later isolated from infected pea leaves (Isogai *et al.*, 1992; Sano *et al.*, 1993). Furthermore, MDV was shown to be serologically related to faba bean necrotic yellows virus (FBNYV), an isometric ssDNA virus occurring in North Africa and western Asia (Katul *et al.*, 1993; Franz *et al.*, 1996). Thus, MDV is now considered to be a member of the genus *Nanovirus*, which includes FBNYV, subterranean clover stunt virus (SCSV; Chu & Helms, 1988), banana bunchy top virus (BBTV; Harding *et al.*, 1991) and coconut foliar decay virus (CFDV; Randles & Hanold, 1989).

To obtain basic information on the ssDNA genome of MDV, and to define its taxonomic relationship to other nanoviruses, we cloned and sequenced MDV DNA. Here, we report the sequences of ten ssDNA components of MDV, and compare them with those available for related viruses, together with the sequences of four newly identified DNA components of FBNYV (C7–C10) reported in the preceding paper (Katul *et al.*, 1998).

Author for correspondence: Yoshitaka Sano.
Fax +81 75 724 7764. e-mail ysano@ipc.kit.ac.jp

The sequence data reported in this paper will appear in DDBJ under accession numbers AB000920–AB000927 (C1–C8), AB009046 (C9) and AB009047 (C10).

Methods

■ **Virus and DNA.** The N isolate of MDV, obtained from naturally infected broad bean, was maintained in broad bean plants, and propagated in pea plants using *A. craccivora* as a vector. Virus was purified from infected pea leaves by the method described for FBNYV (Katul *et al.*,

1993). ssDNA was extracted from purified virions with phenol–chloroform (1:1), and precipitated with 0·1 vol. 3 M sodium acetate, pH 5·2 and 2 vols ethanol.

Replicative-form (RF) dsDNA was isolated from infected pea leaves by a modification of the method of Sunter *et al.* (1984). Leaf material was pulverized in liquid nitrogen and ground in 2 vols 50 mM Tris–HCl, pH 7·8 containing 100 mM NaCl. The homogenate was mixed with 2 vols 0·2 M NaOH–1% SDS and then with 1·5 vols 3 M potassium acetate, pH 4·8. After incubation at 0 °C for 1 h, the mixture was centrifuged at 10000 ***g*** for 10 min. The supernatant was extracted once with phenol–chloroform (1:1), and then with chloroform. The aqueous phase was removed and mixed with 2 vols ethanol, and then centrifuged. The crude nucleic acid pellet was dissolved in TE (10 mM Tris–HCl, pH 8·0; 1 mM EDTA) containing 10 μg/ml RNase A (Pharmacia), incubated at 37 °C for 30 min, and electrophoresed on a 1·2% agarose gel containing 0·5 μg/ml ethidium bromide. The viral dsDNA band was excised from the gel, electroeluted and precipitated with ethanol. The pellet was washed with 70% ethanol and finally resuspended in sterile distilled water.

■ **Construction and sequence analysis of DNA clones.** Recombinant DNA techniques were performed as described by Sambrook *et al.* (1989). In preliminary experiments, RF DNA was cleaved with restriction enzymes with four-base recognition sites, such as *Afa*I, *Hae*III and *Sau*3AI. Digestion products of these enzymes showed multiple partial-length fragments on agarose gels, which gave a total DNA size of greater than 1 kbp, indicating that the MDV genome is comprised of multiple DNA components (not shown). To construct full-length DNA clones, more than 20 restriction enzymes were used. Full-length clones of C1, C2, C3, C5, C6 and C8 were obtained from RF DNA by digestion with *Sph*I, *Xba*I, *Kpn*I, *Sal*I, *Sac*I and *Bam*HI (Takara), respectively, whereas full-length clones of C4 and C7 were obtained as *Hin*dIII fragments. Aliquots of RF DNA (1 μg) were digested with the above enzymes, and the resulting fragments were ligated into the corresponding sites of pUC19. Recombinant plasmids were transformed into competent *Escherichia coli* JM109 (Takara).

Full-length cDNAs of C9 and C10 were synthesized by PCR with the following back-to-back primers: C9(+), 5′ TAATGTAATGAAGAACACTA 3′; C9(−), 5′ CAGTTCAATATACACTCTAT 3′; C10(+), 5′ CATAGATGGACCTTGGGAG 3′; and C10(−), 5′ GCGGTTTCTTTCTTCTGGC 3′. The primer pairs were derived from sequences of two partial-length clones from a random library containing *Sau*3AI-cleaved RF DNA fragments. The sequences of the two DNA inserts did not correspond to any of the eight components C1–C8. PCR was carried out using either C9(+) and C9(−) or C10(+) and C10(−) with *EX Taq* polymerase (Takara) and viral ssDNA as the template. Amplified 1 kbp fragments were subsequently cloned into *Sma*I-cut, ddTTP-tailed pBluescript SK(+) (Holton & Graham, 1991).

Nucleotide sequences were determined by the dideoxy chain termination method using an ABI 373A automated DNA sequencer. Sequences were determined fully in both directions by generating nested deletions (Kilo sequence deletion kit, Takara), or by primer extension. All sequences were obtained from at least two independent clones. Additional PCR clones covering the unique restriction sites of components C1–C8 were generated using primer pairs flanking the gaps. Junctions of the restriction sites were verified from these partial clones. Data were analysed and compared using the GENETYX-WIN program (Software Development). Database searches were performed in GenBank/EMBL, SWISS-PROT, and PIR using FASTA (Pearson & Lipman, 1988) and BLAST (Altschul *et al.*, 1990).

For sequence comparison, the following sequences were retrieved from GenBank: SCSV (U16730–U16736), FBNYV (X80879, Y11405–Y11409), BBTV (S56276, L41574–L41578, L32166, L32167), CFDV (M29963), maize streak virus (MSV) (X01633), beet curly top virus (BCTV) (X04144), bean golden mosaic virus (BGMV) (M10070, M10080) and porcine circovirus (PCV) (U49186).

■ **Analysis of virus-sense polarity.** The full-length C6 fragment was cleaved and subcloned in both orientations into the *Sac*I site of pBluescript SK(+). The resulting plasmids were cut at the *Sac*II site. Strand-specific RNAs were transcribed from the linearized DNA templates with T3 RNA polymerase (RNA Transcription kit, Stratagene), and labelled by using the ECL direct nucleic acid labelling system (Amersham). Aliquots of DNA samples extracted from purified MDV particles, infected pea leaves and viruliferous *A. craccivora* (15 mg tissue) were separated on 1·2% agarose gels, and transferred onto Hybond N+ membrane (Amersham) in 0·4 M NaOH. Hybridizations were carried out according to the supplier's protocol.

Results and Discussion

Characterization of ten ssDNA components

Ten circular ssDNA components (C1–C10) associated with the MDV N isolate were identified and sequenced. Analysis of the sequences revealed that each DNA contained only one potential open reading frame (ORF) that encoded a protein larger than 10 kDa. The sizes, putative motifs and functions of the single ORFs, together with possible gene regulation signals, are shown in Fig. 1. All the DNA components had a region of 29–34 bases capable of forming a stem–loop structure, as has been identified consistently in the genome components of other nanoviruses, as well as in those of geminiviruses and PCV (Meehan *et al.*, 1997). Furthermore, C4–C9 also had certain similar sequences in the non-coding regions. The lengths of the common regions varied from 47 bases in C8 to 437 bases in C4, C5 and C7, in which more than 93% of the nucleotides were identical. C1, C2, C3 and C10 did not contain the equivalent regions due to the large sizes of their coding regions. Analysis of the predicted amino acid sequences revealed that C1, C2, C3 and C10 all encoded putative replication-associated (Rep) proteins. The major ORFs of these components were similar in size (281–284 amino acids), and each contained a NTP-binding motif (Gorbalenya *et al.*, 1990) as well as three other motifs typical of replication-initiator proteins involved in rolling-circle DNA replication (RCR-I, II and III; Koonin & Ilyina, 1992) (Fig. 1). However, they had only limited amino acid identity (42–58%) to each other.

Potential TATA boxes and AATAAA-like polyadenylation signals were appropriately located for all the major ORFs of the ten DNA components. Although C3 and C10 did not contain well-defined TATA boxes at appropriate positions, AATAAATA analogues were located 59 and 57 bases, respectively, upstream of the start codons of the predicted ORFs (Fig. 1, marked by grey triangles). All of the proposed poly(A) signals were preceded 19–41 bases upstream by at least one (A/T)TGTA motif (Sanfaçon, 1994; Rothnie *et al.*, 1994), and were followed downstream by a GT-rich stretch (Gil & Proudfoot, 1984; Conway & Wickens, 1985) containing

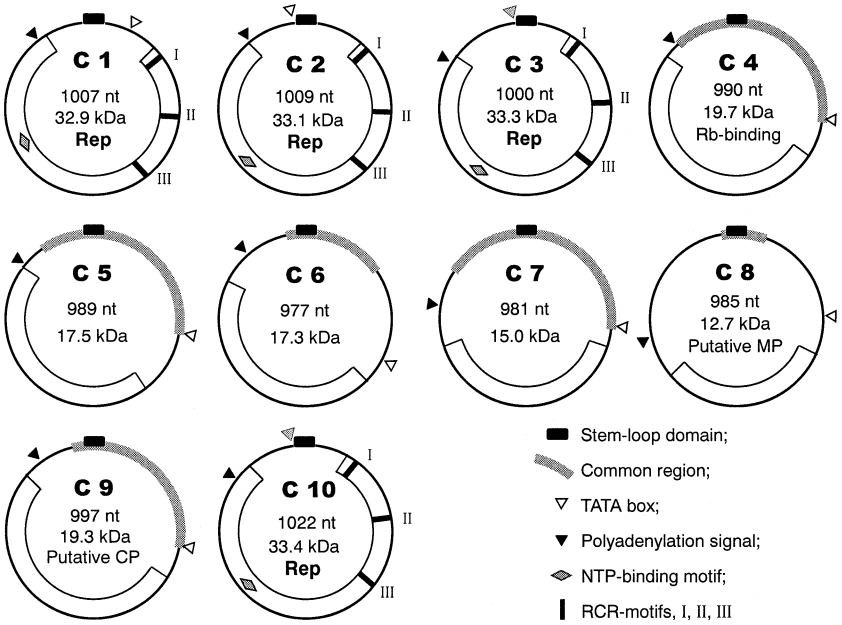

Fig. 1. Diagrammatic representation of the proposed organization of the ssDNA genome of MDV. The single ORF (open box) on the virus-sense strand of each DNA component is shown clockwise.

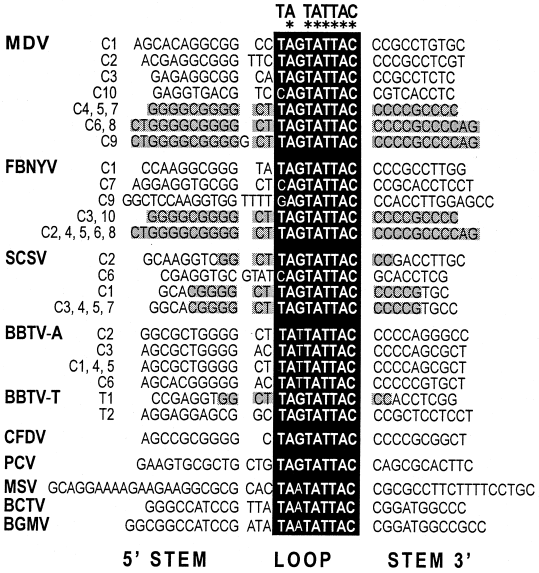

Fig. 2. Alignment of the nucleotide sequences of the putative stem–loop structures of the genome components of MDV, FBNYV, SCSV, BBTV-A (Australian isolate), BBTV-T (Taiwanese isolate), CFDV, PCV, MSV, BCTV and BGMV. The highly conserved nonanucleotide sequences of the loop-forming domain are boxed; asterisks indicate the nucleotides conserved in all the components. The related stem-forming sequences of six MDV components and those of FBNYV, SCSV and BBTV-T are shaded.

the TTG sequence within 47 bases. In addition to the AATAAA signal, the (A/T)TGTA element and the downstream GT-rich sequence have been shown to be involved in correct 3′-end processing of CFDV Rep gene transcripts (Merits *et al.*, 1995).

Fig. 2 shows an alignment of the putative stem–loop sequences of the ten MDV components, together with those identified in the ssDNA components of FBNYV (Katul *et al.*, 1995, 1997, 1998), SCSV (Boevink *et al.*, 1995), BBTV (Harding *et al.*, 1993; Wu *et al.*, 1994; Burns *et al.*, 1995), CFDV (Rohde *et al.*, 1990), PCV and three representatives of the geminivirus genera, MSV, BCTV and BGMV. In addition to the highly conserved loop sequences, particular components of MDV and other nanoviruses had extended sequence similarity within their stem-forming domains (Fig. 2, indicated by grey shading). Most evidently, six MDV components (C4–C9) and seven FBNYV components (C2–C6, C8 and C10) were identical within the stem–loop sequence over at least 30 bases. In contrast, the sequences of MDV-C1, C2, C3 and C10, all of which encoded putative Rep proteins, were less well conserved and had many nucleotide substitutions in the stems and in the 5′-proximal regions of the loop domains. It has been demonstrated that, upon initiation of rolling-circle DNA replication, the geminivirus Rep protein cleaves the RF templates and introduces a nick within a conserved nonanucleotide (5′ TAATATT↓AC 3′) on the virus-sense strand (Heyraud *et al.*, 1993; Laufs *et al.*, 1995; Stanley, 1995). The Rep protein of BBTV-C1 was also shown to cleave ssDNA *in vitro* between positions +7 and +8 of a nonanucleotide (5′ TATTATT↓AC 3′) (Hafner *et al.*, 1997). The conserved nonanucleotides of the MDV components are thus likely to be involved in initiation of virus DNA synthesis.

Analysis of virus-sense polarity

The polarity of the viral ssDNA was examined with strand-specific RNA probes transcribed from a full-length cDNA clone of MDV-C6. The complementary-sense RNA hybridized with virus-sense DNA present in crude RF DNA, purified

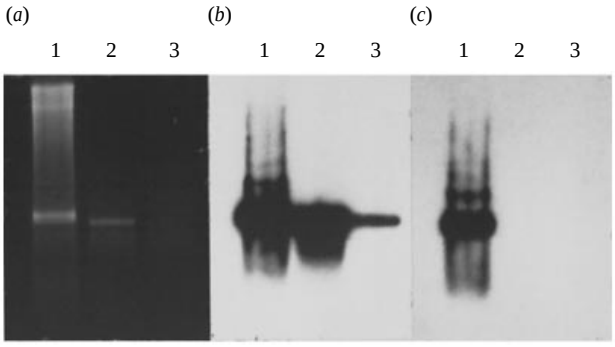

Fig. 3. Analysis of virus-sense polarity using full-length RNA transcripts of MDV-C6. (*a*) A crude RF DNA preparation (1), viral ssDNA (2) and a DNA extract from viruliferous aphids (3, DNA bands are not visible) were electrophoresed on agarose gels. After transferring to Hybond N+ membranes, the blots were probed with complementary- (*b*) or virus-sense (*c*) RNA transcripts.

virions and viruliferous aphids (Fig. 3*b*, lanes 1, 2 and 3, respectively), whereas the virus-sense probe hybridized only with the RF DNA sample (Fig. 3*c*, lane 1), indicating that only the coding strand was encapsidated in particles. The results also indicate that MDV does not multiply in the aphid vector, since complementary-sense DNA could not be detected in viruliferous aphids. The polarities of other components, as shown in Fig. 1, were deduced by analogy to C6: all the MDV components had a conserved nonanucleotide within the putative loop domain (Fig. 2), and six, including C6, also shared sequences surrounding the stem–loop domains.

Predicted genome organization and relationships with other nanoviruses

The sequences of the putative genome products encoded by the major ORFs of MDV-C1 to -C10 were aligned and compared with those identified in the DNA components of FBNYV, SCSV, BBTV and CFDV. Calculated amino acid identities of the putative Rep and non-Rep proteins are shown in Tables 1 and 2, respectively. Three of the four putative MDV Rep proteins showed striking similarity to particular Rep proteins of FBNYV and SCSV: the Rep proteins encoded by MDV-C10, C2 and C3 had 89·8, 82·3 and 73·0% amino acid identity to those encoded by FBNYV-C7, SCSV-C2 and SCSV-C6, respectively (Table 1). Comparisons also indicated at least 35·7% amino acid identity between the putative Rep proteins of MDV and other nanoviruses. When the sequences were compared with the putative Rep protein of PCV (Meehan *et al*., 1997) or with those of MSV, BCTV and BGMV, the four MDV Rep proteins showed only weak similarity (17·7–24·5% amino acid identity, not shown).

In addition to the Rep proteins, both MDV and FBNYV have six putative gene products, whereas only five and four putative non-Rep proteins have been identified for SCSV and BBTV, respectively. MDV-C7 and FBNYV-C6 each encode a protein with a molecular mass of 15 kDa, which has no counterpart in the putative proteins of SCSV and BBTV. Likewise, a homologue of the proteins encoded by FBNYV-C3 and SCSV-C7, which are similar to that encoded by MDV-C5, has not been found in the BBTV components sequenced to date. While six DNA components have been identified in an Australian isolate of BBTV, the one named BBTV-C2 does not contain a relevant ORF (Burns *et al*., 1995). In addition, BBTV-C1 has a second, small ORF within the major ORF (Beetham *et al*., 1997). However, we did not find an equivalent small ORF in the MDV components. The proposed CFDV genome differs from other nanoviruses in that the single ssDNA component of 1291 bases potentially encodes six ORFs, which are thought to be transcribed bi-directionally (Rohde *et al*., 1990). Comparison between CFDV and MDV of all putative proteins did not show

Table 1. Comparison of the four putative Rep proteins of MDV with those of related ssDNA viruses

Percentage amino acid identities calculated for putative Rep proteins are shown. Numbers in parentheses are the overall nucleotide identities of the DNA components. Values exceeding 70% are shown in bold.

	Percentage identity												
	MDV			FBNYV				SCSV		BBTV			
	C2	C3	C10	C1	C2	C7	C9	C2	C6	C1	T1	T2	CFDV
MDV C1	42·1 (56·7)	49·3 (59·3)	56·0 (60·9)	52·8 (61·8)	41·3 (55·1)	54·9 (61·9)	49·7 (59·6)	40·6 (55·6)	47·9 (56·6)	40·7	44·0	41·8	36·6
MDV C2	–	43·9 (56·6)	44·4 (56·3)	62·4 (67·7)	35·7 (51·4)	45·1 (57·4)	41·0 (54·0)	**82·3** (**75·7**)	45·2 (57·3)	36·1	53·6	53·5	44·5
MDV C3	–	–	57·9 (63·3)	48·1 (58·5)	39·7 (53·4)	59·0 (63·2)	51·0 (58·3)	41·1 (55·8)	**73·0** (**72·5**)	37·5	39·8	48·8	36·5
MDV C10	–	–	–	49·3 (58·7)	39·7 (53·7)	**89·8** (**86·2**)	47·9 (60·0)	43·7 (54·6)	57·0 (61·7)	37·4	41·0	49·3	41·5

Table 2. Comparison of putative MDV non-Rep proteins with those of FBNYV, SCSV and BBTV

Percentage amino acid identities calculated between related non-Rep proteins (see text) are shown. Numbers in parentheses are the overall nucleotide identities between the corresponding DNA components. Values exceeding 70% are shown in bold.

	Percentage identity to MDV					
Virus	C9 19 kDa (CP)	C8 13 kDa	C4 19 kDa	C5 17 kDa	C6 17 kDa	C7 15 kDa
FBNYV	**83·1** (**78·9**)	**76·3** (66·9)	**70·4** (**79·1**)	**72·3** (**78·0**)	**90·9** (**78·3**)	52·0 (60·1)
SCSV	55·5 (59·3)	48·3 (52·2)	44·8 (56·4)	45·4 (58·6)	66·7 (61·0)	–
BBTV	23·7	21·8	21·1	–	46·8	–

any apparent similarity, other than between the Rep proteins (data not shown).

The six putative proteins of MDV encoded by C4–C9 were most identical to the related proteins of FBNYV (52·0–90·9%, Table 2). The percentage amino acid identities calculated between the corresponding five proteins of MDV and SCSV were intermediate (44·8–66·7%), and similar to the levels of identity calculated between the corresponding five non-Rep proteins of SCSV and FBNYV (43·0–66·0%, not shown). In contrast, the amino acid sequence identities calculated between the corresponding four non-Rep proteins of MDV and BBTV were considerably lower (21·1–46·8%). Of the six putative non-Rep proteins, that encoded by MDV-C6 and the related proteins encoded by FBNYV-C8, SCSV-C4 and BBTV-C6 showed higher levels of amino acid identity than other proteins; for example, those encoded by MDV-C6 and FBNYV-C8 were 90·9% identical. In contrast, the proteins encoded by MDV-C7 and FBNYV-C6, which have been found in only these two viruses, shared only 52·0% identity. The ten DNA components identified in this study may represent the entire genome of the N isolate of MDV. The predicted genome organization of MDV is in agreement with that of FBNYV, and we did not detect the presence of additional components. However, due to the small size of the DNA components, it cannot be excluded that other genome components had no appropriate restriction sites, and were therefore missed.

The putative protein encoded by MDV-C9 had a molecular mass of 19·3 kDa, and showed 83·1, 55·5 and 23·7% amino acid identity to the capsid proteins (CPs) encoded by FBNYV-C5 (Katul *et al.*, 1997), SCSV-C5 (Chu *et al.*, 1993) and BBTV-C3 (Wanitchakorn *et al.*, 1997), respectively. From this result, and the observation that MDV and FBNYV share common epitopes (Franz *et al.*, 1996), it is likely that the 19·3 kDa protein encoded by MDV-C9 is the virus CP.

The putative 12·7 kDa protein encoded by MDV-C8 had 76·3, 48·3 and 21·8% amino acid identity to those encoded by FBNYV-C4, SCSV-C1 and BBTV-C4, respectively. Based on structural similarity to the movement protein (MP) of MSV (Boulton *et al.*, 1993; Dickinson *et al.*, 1996), these 13 kDa proteins have been suggested to be movement proteins (Burns *et al.*, 1995; Katul *et al.*, 1997). Although none of these proteins has significant sequence similarity to the MSV MP, they all have a hydrophobic domain of 25–30 amino acids in the N-terminal region (not shown), which is also present in the MSV MP.

Database searches with the amino acid sequences of the putative MDV proteins revealed no significant matches other than with related nanovirus proteins, and the functions of the proteins encoded by MDV-C4, C5, C6 and C7 remain unknown. However, the 19 kDa proteins encoded by MDV-C4, FBNYV-C10 (sequence not shown), SCSV-C3 and BBTV-C5 all contain the consensus retinoblastoma (Rb)-binding motif (LXCXE) at equivalent positions, suggesting that they may be involved in controlling the host cell cycle (Fig. 4*a*, boxed). No other Rep or non-Rep protein of MDV, FBNYV, SCSV or BBTV contains the LXCXE sequence. The Rb tumour suppressor is the key regulatory factor of cell cycle progression at the G1/S boundary. Tumour virus oncoproteins are known to inactivate Rb by forming a stable complex through the LXCXE motif, thereby driving the host cell cycle into S phase, where the cellular environment is suitable for replication and transcription of viral DNA. The Rep protein of wheat dwarf virus (WDV), a monopartite geminivirus, has been shown to interact with the human (Xie *et al.*, 1995; Collin *et al.*, 1996) and maize Rb proteins (Xie *et al.*, 1996) through its LXCXE motif. The proposed LXCXE motifs are surrounded by many acidic amino acids (Fig. 4*b*, shown in bold), as is commonly observed in viral and cellular proteins which interact with Rb. Also, the 19 kDa proteins each contain a basic amino acid cluster (R or K) in the N-terminal domain (Fig. 4*a*, boxed), which closely resembles the bipartite motif, a nuclear targeting signal (rrXXXXXXXXXXrrrrr, where r is either R or K; Dingwall & Laskey, 1991). The presence of a nuclear targeting signal is also a common feature of tumour virus oncoproteins, and is

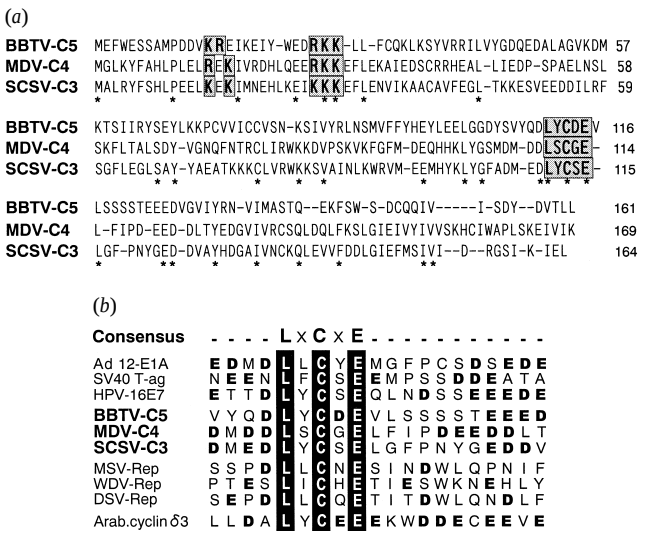

Fig. 4. (*a*) Amino acid sequence alignment of the putative proteins encoded by BBTV-C5, MDV-C4 and SCSV-C3. Identical amino acids are shown by asterisks. The numbering of the amino acids is given on the right. Amino acids constituting putative motifs (see text) are marked by shaded boxes. (*b*) Comparison of the LXCXE motif in the 19 kDa proteins of MDV-C4, BBTV-C5 and SCSV-C3, with those of animal DNA virus oncoproteins, geminivirus Rep proteins and plant D-type cyclin (Soni *et al.*, 1995). The key residues of the motif (L, C and E) are shown in white on black and acidic residues (E and D) frequently occurring around the motif are shown in bold. Abbreviations and sequence sources (accession no.) are: adenovirus type 12 (Ad 12) E1A protein (X73487), SV 40 large T antigen (T-ag) (J02400), human papilloma virus type 16 (HPV-16) E7 protein (SWISS-PROT P03129), maize streak virus (MSV) (X01633) wheat dwarf virus (WDV) (X02869), Digitaria streak virus (DSV) (M23022) and *Arabidopsis thaliana* cyclin δ3 (X83371).

consistent with the general observation that the Rb family are nuclear proteins (Ach *et al.*, 1997). Hence, these facts may indicate that one of the essential functions of geminivirus Rep proteins is undertaken by the independent small proteins of MDV, FBNYV, SCSV and BBTV.

Association of multiple Rep components seems to be a common feature of nanovirus genomes. So far, two distinct Rep protein-encoding components have been identified from an isolate of SCSV (Boevink *et al.*, 1995) and from a Taiwanese isolate of BBTV (Wu *et al.*, 1994). The possible origins of the apparently interchangeable Rep protein-encoding components, as well as their biological significance, are not known. The MDV isolate used has been kept for over 15 years, indicating that the four Rep components are maintained stably. However, there is no evidence that all four Rep proteins are essential for virus replication. Just one of the four Rep components may suffice to support virus replication, and the other three may be dispensable. In the case of BBTV, one Rep protein-encoding component, C1, has been consistently identified in BBTV isolates from different geographical regions, and is thought to represent an integral part of the BBTV genome (Karan *et al.*, 1994), and some additional Rep components have been suggested to be satellite DNAs (Horser *et al.*, 1996; J. L. Dale, personal communication). Two of the MDV Rep components, C3 and C10, had an imperfect TATA box (AATAAATA) at the expected positions, which suggests that their putative Rep genes are inactive or transcribed at a reduced rate. To verify this, it will be necessary to examine viral transcripts by Northern hybridization or by RACE analysis. Furthermore, infectivity studies using the cloned DNAs will be required to ascertain the total number of genome components essential for virus replication.

The observation that there are very similar counterparts of the putative Rep components from MDV, FBNYV and SCSV (Table 2) indicates that the three viruses have evolved from a common origin, which was potentially associated with multiple, distinct Rep components. Since MDV, FBNYV and SCSV have the same aphid vector and all infect legumes, the diversity of their Rep components presumably resulted from recombination events. Based on the overall sequence similarities, MDV and FBNYV are more closely related to each other than to SCSV. SCSV was first reported in the early 1950s as a serious pathogen of subterranean clover (*Trifolium subterraneum* L.), and subsequently spread throughout Australia (Chu *et al.*, 1995). The relatively low sequence similarity of SCSV to the other two viruses may reflect the geographical or evolutionary isolation of SCSV from MDV and FBNYV.

Analysis of the sequence variability of BBTV-C1 and -C6 reveals that BBTV isolates from more than eight different countries can be separated into two large groups, the South Pacific and the Asian groups (Karan *et al.*, 1994, 1997). The sequence variability between the two groups is 9·6% (nucleotide) and 5·6% (amino acid) in C1, and 14·5% (nucleotide) and 6·7% (amino acid) in C6. Results from a previous serological study led to the suggestion that MDV and FBNYV are strains of the same virus (Franz *et al.*, 1996). The putative MDV CP encoded by C9 had 78·9% nucleotide and 83·1% amino acid identity to FBNYV-C5. We feel that these levels of similarity are too low for MDV and FBNYV to be different strains of the same virus. Furthermore, none of the corresponding DNA components of MDV and FBNYV have nucleotide sequence identities over 90%. Hence, based on the overall nucleotide and amino acid sequence similarities, we propose that MDV and FBNYV should be regarded as separate species. Similar taxonomic criteria have been proposed for classifying the strains and species of geminiviruses (Padidam *et al.*, 1995; Hong & Harrison, 1995). A certain amount of serological or genomic variability has been identified within isolates of SCSV (Chu *et al.*, 1995) and FBNYV (Franz *et al.*, 1996). For a more accurate definition of strains and species, it will be necessary to examine the sequence variability among different isolates of MDV, FBNYV and SCSV.

We wish to thank Drs L. Katul and H. J. Vetten for supplying sequence information prior to publication. We also thank Dr J. L. Dale for valuable suggestions, T. Fujita for his technical assistance, and the late Dr N. Ogasawara (Plant Biological Defence System Lab) for providing DNA sequencing facilities. This work was supported in part by a grant-in-aid from the Ministry of Education, Science, Sports and Culture, Japan.

References

Ach, R. A., Durfee, T., Miller, A. B., Taranto, P., Hanley-Bowdoin, L., Zambryski, P. C. & Gruissem, W. (1997). RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein. *Molecular and Cellular Biology* **17**, 5077–5086.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Beetham, P. R., Hafner, G. J., Harding, R. M. & Dale, J. L. (1997). Two mRNAs are transcribed from banana bunchy top virus DNA-1. *Journal of General Virology* **78**, 229–236.

Boevink, P., Chu, P. W. & Keese, P. (1995). Sequence of subterranean clover stunt virus DNA: affinities with the geminiviruses. *Virology* **207**, 354–361.

Boulton, M. I., Pallaghy, C. K., Chatani, M., MacFarlane, S. & Davies, J. W. (1993). Replication of maize streak virus mutants in maize protoplasts: evidence for a movement protein. *Virology* **192**, 85–93.

Burns, T. M., Harding, R. M. & Dale, J. L. (1995). The genome organization of banana bunchy top virus: analysis of six ssDNA components. *Journal of General Virology* **76**, 1471–1482.

Chu, P. W. G. & Helms, K. (1988). Novel virus-like particles containing circular single-stranded DNA associated with subterranean clover stunt disease. *Virology* **167**, 38–49.

Chu, P. W., Keese, P., Qiu, B. S., Waterhouse, P. M. & Gerlach, W. L. (1993). Putative full-length clones of the genomic DNA segments of subterranean clover stunt virus and identification of the segment coding for the viral coat protein. *Virus Research* **27**, 161–171.

Chu, P. W. G., Boevink, P., Surin, B., Larkin, P., Keese, P. & Waterhouse, P. M. (1995). Non-geminated single-stranded DNA plant viruses. In *Pathogenesis and Host Specificity in Plant Diseases*, vol. 3, *Viruses and Viroids*, pp. 311–341. Edited by R. P. Singh, U. S. Singh & K. Kohmoto. Oxford: Pergamon.

Collin, S., Fernández-Lobato, M., Gooding, P. S., Mullineaux, P. M. & Fenoll, C. (1996). The two nonstructural proteins from wheat dwarf virus involved in viral gene expression and replication are retinoblastoma-binding proteins. *Virology* **219**, 324–329.

Conway, L. & Wickens, M. (1985). A sequence downstream of A-A-U-A-A-A is required for formation of simian virus 40 late mRNA 3′ termini in frog oocyte. *Proceedings of the National Academy of Sciences, USA* **82**, 3949–3953.

Dickinson, V. J., Halder, J. & Woolston, C. J. (1996). The product of maize streak virus ORF V1 is associated with secondary plasmodesmata and is first detected with the onset of viral lesions. *Virology* **220**, 51–59.

Dingwall, C. & Laskey, R. A. (1991). Nuclear targeting sequence – a consensus? *Trends in Biochemical Sciences* **16**, 478–481.

Franz, A., Makkouk, K. M., Katul, L. & Vetten, H. J. (1996). Monoclonal antibodies for the detection and differentiation of faba bean necrotic yellows virus isolates. *Annals of Applied Biology* **128**, 255–268.

Gil, A. & Proudfoot, N. J. (1984). A sequence downstream of AAUAAA is required for rabbit β-globin mRNA 3′-end formation. *Nature* **312**, 473–474.

Gorbalenya, A. E., Koonin, E. V. & Wolf, Y. I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes of small DNA and RNA viruses. *FEBS Letters* **262**, 145–148.

Hafner, G. J., Stafford, M. R., Wolter, L. C., Harding, R. M. & Dale, J. L. (1997). Nicking and joining activity of banana bunchy top virus replication protein *in vitro*. *Journal of General Virology* **78**, 1795–1799.

Harding, R. M., Burns, T. M. & Dale, J. L. (1991). Virus-like particles associated with banana bunchy top disease contain small single-stranded DNA. *Journal of General Virology* **72**, 225–230.

Harding, R. M., Burns, T. M., Hafner, G., Dietzgen, R. G. & Dale, J. L. (1993). Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. *Journal of General Virology* **74**, 323–328.

Heyraud, F., Matzeit, V., Kammann, M., Schaefer, S., Schell, J. & Gronenborn, B. (1993). Identification of the initiation sequence for viral-strand DNA synthesis of wheat dwarf virus. *EMBO Journal* **12**, 4445–4452.

Holton, T. A. & Graham, M. W. (1991). A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Research* **19**, 1156.

Hong, Y. G. & Harrison, B. D. (1995). Nucleotide sequences from tomato leaf curl viruses from different countries: evidence for three geographically separate branches in evolution of the coat protein of whitefly-transmitted geminiviruses. *Journal of General Virology* **76**, 2043–2049.

Horser, C., Karan, M., Harding, R. M. & Dale, J. L. (1996). Possible satellite DNAs associated with banana bunchy top virus. *Xth International Congress of Virology*, Jerusalem. Abstract PW52-2, p. 237.

Inouye, T., Inouye, N. & Mitsuhata, K. (1968). Yellow dwarf of pea and broad bean caused by milk-vetch dwarf virus. *Annals of the Phytopathological Society of Japan* **34**, 28–35 (in Japanese with English abstract).

Isogai, M., Sano, Y. & Kojima, M. (1992). Identification of the unique DNA species in the milk-vetch dwarf virus-infected leaves. *Annals of the Phytopathological Society of Japan* **58**, 631–632 (Japanese Abstract).

Karan, M., Harding, R. M. & Dale, J. L. (1994). Evidence for two groups of banana bunchy top virus isolates. *Journal of General Virology* **75**, 3541–3546.

Karan, M., Harding, R. M. & Dale, J. L. (1997). Association of banana bunchy top virus DNA components 2 to 6 with banana bunchy top disease. *Molecular Plant Pathology On-Line* http://www.bspp.org.uk/mppl/1997/0624karan.

Katul, L., Vetten, H. J., Maiss, E., Makkouk, K. M., Lesemann, D.-E. & Casper, R. (1993). Characterization and serology of virus-like particles associated with faba bean necrotic yellows. *Annals of Applied Biology* **123**, 629–647.

Katul, L., Maiss, E. & Vetten, H. J. (1995). Sequence analysis of a faba bean necrotic yellows virus DNA component containing a putative replicase gene. *Journal of General Virology* **76**, 475–479.

Katul, L., Maiss, E., Morozov, S. Y. & Vetten, H. J. (1997). Analysis of six DNA components of the faba bean necrotic yellows virus genome and their structural affinity to related plant virus genomes. *Virology* **233**, 247–259.

Katul, L., Timchenko, T., Gronenborn, B. & Vetten, H. J. (1998). Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. *Journal of General Virology* **79**, 3101–3109.

Koonin, E. V. & Ilyina, T. V. (1992). Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *Journal of General Virology* **73**, 2763–2766.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J. & Gronenborn, B. (1995). *In vitro* cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceedings of the National Academy of Sciences, USA* **92**, 3879–3883.

Matsuura, Y. (1953). Studies on the dwarf disease of milk vetch (*Astragalus sinicus*). *Annals of the Phytopathological Society of Japan* **17**, 65–68 (in Japanese with English abstract).

Meehan, B. M., Creelan, J. L., McNulty, M. S. & Todd, D. (1997). Sequence of porcine circovirus DNA: affinities with plant circoviruses. *Journal of General Virology* **78**, 221–227.

Merits, A., Zelenina, D. A., Mizenina, O. A., Chernov, B. K. & Morozov, S. Yu. (1995). Poly(A) addition site mapping and polyadenylation signal analysis in a plant circovirus replication-related gene. *Virology* **211**, 345–349.

Ohki, S. T., Doi, Y. & Yora, K. (1975). Small spherical virus particles found in broad bean plants infected with milk-vetch dwarf virus. *Annals of the Phytopathological Society of Japan* **41**, 508–510 (in Japanese with English abstract).

Padidam, M., Beachy, R. N. & Fauquet, C. M. (1995). Classification and identification of geminiviruses using sequence comparisons. *Journal of General Virology* **76**, 249–263.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences, USA* **85**, 2444–2448.

Randles, J. W. & Hanold, D. (1989). Coconut foliar decay virus particles are 20-nm icosahedra. *Intervirology* **30**, 177–180.

Rohde, W., Randles, J. W., Langridge, P. & Hanold, D. (1990). Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. *Virology* **176**, 648–651.

Rothnie, H. M., Reid, J. & Hohn, T. (1994). The contribution of AAUAAA and the upstream element UUUGUA to the efficiency of mRNA 3′-end formation in plants. *EMBO Journal* **13**, 2200–2210.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanfaçon, H. (1994). Analysis of figwort mosaic virus (plant para-retrovirus) polyadenylation signal. *Virology* **198**, 39–49.

Sano, Y., Isogai, M., Satoh, S. & Kojima, M. (1993). Small virus-like particles containing single-stranded DNAs associated with milk-vetch dwarf disease in Japan. *6th International Congress of Plant Pathology*, Montreal. Abstract no. 17.1.27 p. 305.

Soni, R., Carmichael, J. P., Shah, Z. H. & Murray, A. H. (1995). A family of cyclin D homologs from plants differentially controlled by growth regulators and containing the conserved retinoblastoma protein interaction motif. *Plant Cell* **7**, 85–103.

Stanley, J. (1995). Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs within the conserved nonanucleotide motif during the initiation of rolling-circle DNA replication. *Virology* **206**, 707–712.

Sunter, G., Coutts, R. H. A. & Buck, K. W. (1984). Negatively supercoiled DNA from plants infected with a single-stranded DNA virus. *Biochemical and Biophysical Research Communications* **118**, 747–752.

Wanitchakorn, R., Harding, R. M. & Dale, J. L. (1997). Banana bunchy top virus DNA-3 encodes the viral coat protein. *Archives of Virology* **142**, 1673–1680.

Wu, R. Y., You, L. R. & Soong, T. S. (1994). Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus. *Phytopathology* **84**, 952–958.

Xie, Q., Suárez-López, P. & Gutiérrez, C. (1995). Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: requirement for efficient viral DNA replication. *EMBO Journal* **14**, 4073–4082.

Xie, Q., Sanz-Burgos, A. P., Hannon, G. J. & Gutiérrez, C. (1996). Plant cells contain a novel member of the retinoblastoma family of growth regulatory proteins. *EMBO Journal* **15**, 4900–4908.

Received 27 April 1998; Accepted 10 August 1998