

Characterization of cassava vein mosaic virus: a distinct plant pararetrovirus

Lee A. Calvert,^{1*} Manuel D. Ospina¹ and Robert J. Shepherd²

¹Centro International de Agricultura Tropical (CIAT), A. A. 6713, Cali, Colombia and ²Department of Plant Pathology, University of Kentucky, Lexington, KY 40546, USA

Cassava vein mosaic virus (CVMV) was found to be widespread throughout the north-eastern region of Brazil. The complete sequence of CVMV was determined, and the genome was 8158 bp in size. A cytosolic initiator methionine tRNA ($tRNA^{met}_{i}$)-binding site that probably acts as a primer for minus-strand synthesis was present. The genome contained five open reading frames that potentially encode proteins with predicted molecular masses of 186 kDa, 9 kDa, 77 kDa, 24 kDa and 26 kDa. The putative 186 kDa protein had regions with similarity to the zinc finger-like RNA-binding domain that is a common element in the capsid proteins and similarity to the intercellular transport

domain of the plant pararetroviruses. The predicted 77 kDa protein had regions with similarity to aspartic proteases, reverse transcriptase and RNase H of pararetroviruses. This gene order was confirmed by the amplification of similar PCR products from total DNA extracted from CVMV-infected cassava plants. The genomic organization of CVMV was different from the organization of either the caulimoviruses or badnaviruses. In comparisons of the regions with the reverse transcriptase motif, CVMV was grouped between the caulimoviruses and badnaviruses. It appears that CVMV is distinct from the other well-characterized plant pararetroviruses.

Cassava vein mosaic virus (CVMV) has isometric particles 50–60 nm in diameter (Kitajima & Costa, 1966) and the genome is DNA (Lin & Kitajima, 1980). The only known host for CVMV is cassava and the virus can be transmitted by mechanical inoculation. There are dense inclusion bodies in the cytoplasm of cells infected with CVMV (Kitajima & Costa, 1966). Based upon this limited information, CVMV is listed as a tentative member of the caulimovirus group (Hull, 1984).

The two groups of plant viruses that contain dsDNA genomes are the isometric caulimoviruses (reviewed by Guilfoyle *et al.*, 1987; Hull *et al.*, 1987) and the bacilliform badnaviruses (Lockhart, 1990; Medberry *et al.*, 1990). The viruses in both groups share a number of features with other pararetroviruses (Medberry *et al.*, 1990; Temin, 1989; Toh *et al.*, 1983). Some of the common elements of pararetroviruses include a virally encoded replicase (Mason *et al.*, 1987) that has reverse transcriptase and RNase H activity. For the well-

characterized plant pararetroviruses, the primer for minus-strand replication is cytosolic initiator methionine tRNA ($tRNA^{met}_{i}$).

Cassava plants infected with CVMV display a range of symptoms that includes a chlorosis that follows the veins and can coalesce to form a mosaic pattern. There is often leaf distortion and epinasty of the young leaves. Cassava is a vegetatively propagated crop, and CVMV is transmitted readily through the stem cuttings used as reproductive material. When CVMV-infected stem cuttings germinate the plants are often stunted. This is normally followed by a flush of leaves that have no apparent symptoms.

Since there is little information on the distribution or the characterization of CVMV, these studies were undertaken in order to gain insight into the prevalence of the disease caused by CVMV and to better characterize this virus. We conclude that CVMV is a plant pararetrovirus that is distinct from both the badnaviruses and the caulimoviruses that have been described to date.

Surveys to determine the presence of viruses affecting cassava were made in eight states of Brazil. The states were Paraná, Santa Catarina, Para, Amazonas, Ceará, Pernambuco, Alagoas and Bahia. The distinctive symptoms of CVMV were the primary method of determining the range of the disease. The presence of CVMV was

* Author for correspondence. Fax +57 2 445 0273. e-mail CIAT-VIROLOGY@CGNET.COM

The sequence data reported in this paper have been submitted to GenBank/EMBL/DDBJ and assigned the accession number U20341.

TGGTATCAGAGCTTAGTTAAAATAATAATGGATAGTAAAGATTACACAACCTTAATCTAGAGAACATAGTTAAGGTTAACACAAAATTACCATAGTCATATTACAGT
 G G F T Y A N F T P Y I H G D M G F G E K Q L N W T N K L L W N K L G K L N 71
 ATGGAGGATTACTTATGCAAAATTACACCATATTATACATGGAGATGGATTGGAGAACATAACACTGAATTGGACTAAATAATTATGGATAAATTAGGTTAAATTAA
 I K D T Q I L M Q N N I S E E Q G H N T L E I S L E A Q K I A R E N L A D R I N 111
 ATATCAAAGACACACAAAATTCTAACATGGAGAAATAATTAGTAAAGAACACAAATAAGCTTAGAGCACACAAAATTAGCTAGAGAAAATTAGCTAGATAAATTATT
 Q N I N T S I D F K L W K M N K E N L E R Q E L L R Q I N E L K E E I S K L K 151
 TACAAAATATACTAACATGGAGTTAAAGGATAAAAGGAAATTAGAGAAAGGCAAGAATTATTAAGACAAATAATAGGTTAAAGAGAAATAAAATCATTAA
 N I P S T V A I I P T N T Y T I N M I R T E D W K Y F K Y I E K E L V O N 191
 AAAATAACCAAGTACAGCTAACAAATAACCAACTAATACTTAACTATTATGATTAGGACAGAACAGAAGATGGAAATAATTATGAAAGAGATTAGTTCAGAATA
 T E A I A K I L D N S Y I I N D N L L G L Y E R Y E E I N P T P K Y P K R P E T 231
 AAACGTGACCAAGTACGAAATAATTAGTAAATTAGTAAATAATGATAATTAGGTTATTATGAAAGAGATAAGGAAATAATTACCCACCTAACACATATAAAAGCCAGAA
 I F D T P Q Y A K Y I N R N Q K R Q E E Y E K Q Q E L K K E N E N K E Y Q E L E 271
 CTATATTGATACACCTCAATATGCAAAATAATTAGGAACTAAAAAGAACAGAACAGAGATGGAAAGGAAATTAGGAAATAATTACAGATAAATTAGCTAGGAA
 W K E K Q O K D K G K G I Q T V Y P T L I I P D I K P E K O K K E D M M L E M I 311
 AATGCGAAAGAAAACACAAAAGATAAGGTAAGGAAATAACACTGTATACCGACATTAATACAGATAATAACCTGAAAGAACAAAGAGATAATGATGTTAGAAATGA
 N L Q N E L E Q L K I Q R H K E C H K Q A E L T K I Q M L E E E L E E D P 351
 TTAAAATTACAAATAAGGTAAGGAAATAAAATTACAGATAAAAGAACAGAACAGCAGGTTAACATAACAGATGTTAGGAGGAAATTAGGAAAGAATAATTAGTC
 D N L E K V E N L N I Q I Q I S S D I S E S S E I N E I S D N E T Q C I S G 391
 CTGATAATCTAGAAAAGGTTTTAAACATCAAATAACATGACTAGTGTAAAGTACAGATAAATTAGGAAATAACAGATAAAGGAAATTACCGAGAA
 D S D Y N N E Q I N V K I G E E E Y E K Q K D N Y R Y Y K P O P P Y Y K D I R R 431
 GTGAGATGTCATAATAATGACAAATAATTAGTAAAGGAGAAAGATAATGAAATAACAGAAATTATAGATAATAACCCACACCCATATTACAAAAAGATAAGAA
 E R Q Y K G O S S O R A D Y I K N R R E F E S T Y Q A N M N T T I N D S G E E 1320
 GAGAAAGGCAATAAAAGGTCAAGGTCACAAAGAGCAGATTACATAAAACAGAGAACAGAATTGTGATCTACATTCAGCAAAATAATGATAACACATAAGGATGTT
 L N L D C T T P E E A E D R I Q K W T Q S M S I A L V K Q Q L S T H E L M G L R 511
 TTCTAAATTAGTAGGCAACACAGGAGAACAGAGATAAACTAACAAAGTCTGACTCATGTTAGGAAATAACACATAATTCTACAGATAACAGGAA
 R R T F I G N V K E W Y K N L T E N A K Q K L E G N A P L S L T H E L M G L R 1560
 TAAGAGAACTTTAGGAAATAAGGTTAAACATCAAATAACATGACTAGTGTAAAGTACAGATAAATTAGGAAATAACAGATAAAGGAAATTACCGAGAA
 A E F G K L G I E S D V E H E K K T S I A R H K I L Q L Q I C S M D H O N L N 591
 GAGCGAGAATTGCGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAA
 A Y L C E F Q E Y Y Y S A N Y T E A A S E S E N I L N M F Y S K L P E P W G Q Q V L 631
 ATGCATATTATGTAATTGCAAAATAATTAGTCAAAATAACAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAA
 N G Y L S E I K G K N L L D S I G A R M T Y L Q E F I S D K A C T C C C A G A T G G G A C A A C A A G T 1920
 TAAATGGATGATAATGAGGAAATAAGGTTAAATTAGTAAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAA
 R K I Q L S K N L D C S Y E V G P K Y G C K Q I R P H K R K R Y Y K K Y I P K 711
 CTAGAAAATAACATAATCAGAAAGATCTAGGTTAGCTATTAGTAAAGTAAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAA
 R K Y F N K K R Y K K Y Y R P K K F L K R K N P H K A C K C Y N C G E E G H I S 751
 AAAGGAAATAATTAAAGAAAAGATAACAAAATAATTAGCCTAAAGGAAATACTCTAGGCAAAATAAGGAAATAAGGAAATAAGGAAATAAGGAA
 P N C K K P K K K T R I N N L E A L E F K N T E M E N L E F T T N K N D I V W 791
 GTCCAAATTGTAACACAAAAGAAAATAAGGAAATAAAATTAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAA
 E E I V Q P L H Y E E E K Y K G N Y S D R I L Q N P Y Y I N S I S I E L 831
 TAGAAGAAAGGATGTTACACCACTTACATTAGAAGAAGAAAATAACAGGAAATAACTCATGAGATAATTACAGATAAATTACAGATAAAGTATT
 D N L D W E F F E Y Q E D E D D L E Y Q N F V Y Q E S N D N W Y S D Q E N W Y S 871
 TAGATAATTAGGATTTGGAAATTGAAATACAGAGATATTGAAAGTATTGAAAGTATTGAAACATTAGGAAATAAGGAAATAAGGAAATAAGGAA
 D E Q Y L G I Y M F I G E T S G E N N Q D N M E G I I K E Y N K T E P E K I N K 911
 GTGATGACAAATAATTAGGAAATAATGTTAGGAGAACATAGGAAATAACAGGAAATAACAGGAAATAACAGGAAATAACAGGAAATAACAGGAA
 I I F T S E K F Q K I M E D L N M T K D K I F H N N K L K L F G K K E I A Y 951
 AAAATAATTACCTCAGAAAATAACAAATAATTAGGAAATAAGGAAATAACAGGAAATAACAGGAAATAACAGGAAATAACAGGAAATAACAGGAA
 Y I V T D I E H P I D V K Y V Q N O D Q K I I N L P L Y N O E I F E N E E I Q P 991
 ACTATAGTATCAGATGTCAGAACATCTTACGTGAAATAAGGAAATAACAGGAAATAACAGGAAATAACAGGAAATAACAGGAAATAACAGGAA
 D K Q N K N I R N H I A A V E I V V K A Y F R E G I D T P F E I I L C D D R I 1031
 CTGACAAGGATCAAAATAAGGAAATAACATATTAGCAGCTAGGAAATAAGGAAATAACAGGAAATAACAGGAAATAACAGGAAATAACAGGAA
 T Y P Q E G S L V E V L I G N L I Y Q K V K F T K I I N Y S I S I E D K N L D K 1071
 TAACATATCACAGGAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAA
 S L M V Y W N L E G I K M I K D S K I F S I R L R N L Y V L S N K H I V K N K 1111
 AAAGTTGACTATCTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAACTTGTAGGAA
 Q Y N G N I I I E P I F Q D F V I O N Q N N R N Y I E Y G K P G K F D R T K L K S Y 1151
 ACAAAATAATGGAATAATAATTAGGAACTTATTAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAAATAAGGAA
 S R E F P L R L D D R T N I Q R E K Q D K I E K A D H N L E F Q K L E N H L N 1191
 ATAGTAGAAGGATTTATGAACTGAGCTTAAGGAGATGAGAACAAATACTGGAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAAC
 Y Y S Q Q G Q S S N V L D P I P K I L K I E N T C K N N Y K Q F H I I G K I T E G R 1231
 ATTATTATCTAACAGGACAAAGTCTAATTGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAA
 L N K F Y P I L D T G A C A H H S Y I S S K I L E D E K L V S N K L S K V V T S 1271
 GATTAATAATTCTGGCAATTAGGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAA
 N A D N H K I Y D R N T E V I I E L D K N K N E K Y K I F N I G L V D Q L R 1311
 ATAACGCGATAATGAAACATAATTATGACAGAACACAGAGTAATAATTAGATAAAATAATGAAATAATTATGAACTTATGAACTTATGAACTTATGAA
 L E G G K A E I I L G M N I L Q N L P Y C I T D D Y L E I N L G F R C I K R 1351
 TTTAGAGGAGGAAAGCTGAAATAATTAGGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAACTTATGAA
 R I K K D I F E I R E D L Q Q M N V L N E * M S L M N S T N P E F I E Y H T S W K P H K 23
 ATAGAAATAAGGAAAGATAATTGCGAAATTAGGAGAGTTACACAAATGAATGTCCTTAATGAATAGTACAAATTCCAGAAATTATAGATACTCATG
 I E L D N I Y H S Y G Y Y V Y T R S V I K R F E N K H L I K T T Y K R I F S P H 63
 ATAGAATTAGTGTATAATATATCATTCATGATGATTATGTTGATACAGAATGTTAAACATGTTAAACATGTTAAACATGTTAAACATGTTAA
 E N I V L H F R *

Fig. 1. For legend see opposite.

Y L S R Y N S S D T D G R Y D E A N T * M E D M M K Q I L E K L N T I E K N I S E T N I R I E K I 29
 ATTATTTGAGCAGGTACAATTCTGATACTGATGGAGATATGATGAAGAAAATCTGAGAAAACCTCAATACAAATTGAAAAAATTTCTGAAACAAAATATACGAATTGAAAAATA 6360
 E K E Q E L K R K V E L Y G K E P E E K K L H K E N I E K L S S S I E D K I Q N 69
 GAAAAAGAACAAAGAGCTTAAAGAAAAGTAGAGCTCTATGAAAAGAACCCAGAGAAAAAATTCATAAAGAAAATATAGAAAATTAAGCTCTTCTATTGAAAGACAAAATCATACAAAAT 6480
 I D K K L K K I E N V E Q Y Q W K N I V K I N K P L S V G E K Y M E N F K K I 109
 ATTGAATAGAAATTAAAGAATTGAAATGTCGAGAACAAATATCAGTGAAAATATAGCTAAAATTAACCCACTTCTGGGAGAAAATATGGAGAATTTCAAAAGATA 6600
 L V Y L G E B K H P K L E E L Y S L T D Y N K L V A D I Y T D R N L V I S A Y N Y 149
 CTGTTTATCTTGAGAAAACATCTAAATTGGAAGATAATTCTTGACAGATTACACAAAATTAGTAGCAGATATTACTGATGAAATTAGTAATTTCAGCATATAATTAT 6720
 G L L Q V L Y I E H P S Q L E L F D E N I K L A Y M K F R N V T R H N L Y I C E 189 M R I 3
 GGATTAATTCAGTGTCTTATATTGAGCATCGCTCTCAATTAGAGTTATTGATGAAAATATAAAACTGCTTATATGAAATTGTTACAGGCACAACTTATATATGCGAA 6840
 F T V Q W N F M P M I K G * 201
 Y S A M A E P Y D K G V I P K I E I I K F G I T Y S K L K Y D E V Y E H Q P I E 43
 TTTCAGTCAATGGCTGAAACCTATGATAAAGGGTGTATTCCAAAATGGAAATAATTAAATTGGAAATTACTTATTCAAAATTAATGATGAACTATGAACTACAACCAATAG 6960
 R L D L K F K I S Q K R A L G I L V I Q K E I E N L N G N V W L Y S N L D D G R L 83
 AAAACTTGTGTTAAAGAAAATTATATCACAAAAGACCAATTAGGGTACTGTGTTCTAACAAAAGAAAATGAAACTTGTGAACTTGTGACTCTAACCTAGATGAGAC 7080
 I L S N H N K A E N V E V K N I L S D W S K K L S I P E G N Y P C S I K N P M 123
 TTATATTATCAAATCACAAACAAAGCAGAAAATGAGCTCAAAATATATTATCAGATGGCTCACAAAATATCCATTCTCAGAGGTAAATTATCAGATGAGCTAACAGATCCAA 7200
 F T G K T M E V L C E L S K Q O I N M R H I C N L C S K N K V Q I Q D P I L P 163
 TGTTACGGGAAAATCTGAAAGTATATGTCAGCTCACCAAGAAGCAGATCAATATGCCAACATATGCCAACCTATGTCAGGAAATGAGCTAACAGATCCTATACTGC 7320
 E Y E E E Y V E I E K E E P G E E K N L E D V S T D D N N E K K K I R S V I V K 203
 CAGAATACCAAGAAGAAATCAGTAGAAAAGAAGAACCCAGGCCAAGAAAAGAATCTGAAAGACGTAAGGACTGACCAACAAATGAAAAGAAGATAAGCTCGGTGATGTGA 7440
 E T * 205
 AACAGACATAGGGACACATGTAAGCTGAAAATGTAAGGGGAAAGACTTACATCACAAAGGAATCTTATCCCCTACTTATCTCTTATATTTCCCTGTCATTTTGCCCT 7560
 GAGTTTCTCTATAAGGAAACAGTTCGGCATTGTGAAACAAAAGAAAATTTGGCTAAGCTTCTCTGAGACTTCTCTGAGGATAACCTAGAGAAAATTGTAAGTTGTAATG 7680
 TTTTAGTTTATAATAATATGTTATGTTTATAATGAGTAGTGTCTTCTAACTTCTCTAAATATAATGAGTAGTGTCTGAGAACAACTGTTGAT 7800
 ATTCCAGAGCTGAACTGAGCTTACGCTGATGAAATCTTCAATTATGCCCTGATACCTGGCTGTGATATTACTGAGGCCCTTATAGACCATGTCGTTGTATGGCATGAC 7920
 ATCTGTTATGTTAGGGAGAAAAGAGATGATCATTCAGAAAGTTTAAATGAGGAACTCATCTGAGGAAATGAGCTAACATGAGTACATTAATGAGTACATTA 8040
 ATTAGGAAATGTTAAAATATTATTGTTTCACTCTAGTTAATTATCCTTGTGTTTATCCATCTATATTCTTGTGTTATAATTCCGCTTCTTCAACCCCTT 8158

Fig. 1. Nucleotide sequence of CVMV genomic DNA and the predicted amino acid sequences of the five ORFs. The numbering of the genome begins with the putative tRNA^{Met}-binding site. A TATA-like sequence with similarity to the 35S promoter of CaMV is underlined.

confirmed by observation of particles using transmission electron microscopy or using a polymerase chain reaction (PCR) specific for CVMV.

Cassava plants infected with CVMV were most prevalent in the semi-arid region of the north-eastern states of Ceará, Pernambuco and Bahia. It was not unusual to find fields with 50% to 100% of the plants infected with CVMV. Some cassava plants with symptoms of vein mosaic were found in the coastal littoral region of Ceará, Pernambuco, Bahia and Alagoas, but the incidence of virus was much lower than in the semi-arid areas. CVMV-infected cassava was not found in any of the other states.

Plants infected with CVMV were propagated in the greenhouse using stem cuttings. Viral DNA was prepared for cloning by a modification of the mini-preparation procedure for the isolation of the DNA of caulimoviruses (Gardner & Shepherd, 1980). Infected leaves were homogenized in 0.2 M-citrate, pH 6.0, containing 1 M-urea and 2.5% 2-mercaptoethanol (10 ml of buffer/g of tissue). Immediately before use, 4% (w/v) polyvinyl-pyrrolidone was added to the buffer. After homogenization, Triton X-100 was added to a final concentration of 2% (v/v), and the homogenate was stirred for 20 min and then centrifuged at 10000 g for 10 min. The supernatant was filtered through two layers of Miracloth and subjected to ultracentrifugation and further treatment as described in the mini-preparation procedure of Gardner & Shepherd (1980). After some preliminary restriction mapping, the CVMV dsDNA genome was cloned at a unique *Bgl*II restriction enzyme site into a plasmid vector designated pCKIZ (Anza, 1982). This plasmid was found to reproduce poorly in many strains of *Escherichia coli*. The pUC119 (Vieira &

Messing, 1988) vector was modified to include a *Bgl*II site in its multiple cloning site and was designated pJAW141. Subsequently, the cloned DNA of the CVMV virus genome was transferred from pCKIZ into the *Bgl*II site of pJAW141 and the clone was designated pCVMV141.

Whenever possible, subclones of pCVMV141 were produced using appropriate restriction enzymes and used for the sequence reactions. A series of nested deletions were produced using an Exonuclease III digestion technique (Henikoff, 1984). The double-stranded cDNA was sequenced using Sequenase (USB) following the dideoxynucleotide chain-termination method (Sanger *et al.*, 1977). The entire sequence of CVMV was determined on both strands and the sequence data were analysed using Seqaid II (D. D. Rhoads and D. J. Roufa, 1989, Kansas State University, Manhattan, Kan., USA) and DNASIS (Pharmacia).

Total nucleic acids were extracted from dried or fresh plant samples using a nucleic acid extraction method (Dellaporta *et al.*, 1983). Five sets of primers were used to initiate PCR to amplify different regions of the CVMV genome. The PCR was carried out with *Taq* DNA Polymerase (Promega) using a programmable thermal controller (MJ Research). The products were purified using the Magic PCR DNA Purification System (Promega) and analysed in agarose gels. In selected cases the PCR products were cloned into the plasmid Bluescript KS(+) (Stratagene) and the cDNA products were sequenced as described above.

The genome of CVMV consisted of 8158 nucleotides of dsDNA with the potential to encode five proteins and a 736 bp intergenic region (Fig. 1). There was a 19 nucleotide region, TGGTATCAGAGCTTAGTT, that

RNA-binding domain

CVMV	ORF I	739	CKCYNCGEEGHISPNC
CaMV	ORF IV	410	••WI•NI••YANE•
CERV	ORF IV	418	••WV•NI••YANE•
FMV	ORF IV	409	••WI•T••CYANE•
SoyCMV	ORF IV	380	•Q•WL•H••CYANE•
CoYMV	ORF III	879	•••I••Q••YANQ•
ScBV	ORF III	736	•••V••SPD•LMKD•
RTBV	ORF III	776	•••I•QD•N•LANR•

Movement domain

CVMV	ORF I	997	KIRNIHLAAVEIVVKAYFREGIDTPFEIILCDDRI
CaMV	ORF I	123	IMSMV••G••K•LL••Q••N•••IK•A•I•••
CERV	ORF I	115	T•SD••FGAIKVLI••R•••NS•IKMA•I•••
FMV	ORF I	120	••SM••G••K•LLT•Q••Q•••SVKMA•I•••
SoyCMV	ORF I	98	L••YV•ISTLQVLI•ST•LK•L••L•LT•R•N•L
CoYMV	ORF III	136	G•EF•IGVMLVRKQILH•KFAG•MAL•VFR•T•W
ScBV	ORF III	132	GLRY••PGILAVRIQPPLHPDWSGKLVF•VFR•I•D
RTBV	ORF III	109	GKLYY•IGMMA•G••GLH•RK•G•KVM•MFY•DSF

Protease

CVMV	ORF III	26	FDTGANICICKKKVLPE
CaMV	ORF V	44	V••••SL••AS•F•I•E•
CERV	ORF V	33	V•••SSL•MAS•Y•I•E•
FMV	ORF V	53	V••••SL••ASRYII•E•
SoyCMV	ORF V	35	I••••TL•FG•R•ISNNW
CoYMV	ORF III	1219	V••••TA•LIQISAI•EN
ScBV	ORF III	1082	L••••TRSCINQVFIEEK
RTBV	ORF III	985	I•S•STHN•ICPTLI•AS

Reverse transcriptase

CVMV	ORF III	296	YYSKFDCKSGFYHI <20> WKVLPGYHNSPSIF <12> DFIIVYIDDILVFS
CaMV	ORF V	336	IF•S••••WQV <20> •N•V•••LKQA••• <12> K•CC••V•••••
CERV	ORF V	317	I••S••••LWQV <20> •N•V•••LKQA••• <13> KYCC••V•••••
FMV	ORF V	328	IF•S••••WQV <20> •N••••LKQA••• <12> K•CM••V•••I•••
SoyCMV	ORF V	291	WF•SL•A•••YYQL <21> •N••S••LKQA•C•Y <11> •H•LA•••••I•T
CoYMV	ORF III	1498	I••••L••••WQV <20> •L•M••LKNA•A•• <12> K••A•••••••
ScBV	ORF III	1350	IF••••L••••HQV <20> •L•M••LKNA•A•• <12> G••A•••••••
RTBV	ORF III	1274	IF••••L•A•••HHM <20> •N•C••IANA•CA• <11> K•ALL•••••IA•

Ribonuclease H

CVMV	ORF III	512	IIEVDASNNAYG <12> EYLCRYNSGTFK <73> QLIDGKTNIIADYL
CaMV	ORF V	547	••T•••DDYW• <14> •L•••A••S• <72> EH•K•TD•HF••F•
CERV	ORF V	531	V••T•••EEFW• <10> ••I•••A••S• <72> EH•A•TK•VF••F•
FMV	ORF V	540	••T•••DSFW• <11> •L•••S••S• <72> EHLE•VK•VL••C•
SoyCMV	ORF V	508	•V•T•••QHSWS <63> LL••K•V•••T <72> E••KSEN•PFEIR•
CoYMV	ORF III	1711	••T•GCMGTW• <15> •RI•A•A••S•N <72> EH••••H•GL••A•
ScBV	ORF III	1563	••T••CATGW• <15> •QI•••A••K•D <72> EH•K••S•GL••I•
RTBV	ORF III	1486	••T•••EEGW• <15> •KIAG•A••N•G <72> EH•K•NK•FLPNF•

Fig. 2. Comparisons of regions of biological significance between CVMV, CaMV, CERV, FMV, SoyCMV, ScBV and RTBV. The regions are the intercellular transport movement domain, the RNA-binding domain, the protease motif, the reverse transcriptase motif and the RNase H activity domain. The name of the virus, the designation of the ORF and the starting position of the amino acid are indicated before each sequence. The symbol (●) indicates identical amino acids as compared to CVMV. The spacing between amino acid blocks is written in angle brackets.

has 16 bases that were complementary (underlined) to the 3' terminus of plant cytoplasmic tRNA^{met}_i (Sprinzel *et al.*, 1987). To conform with previous studies on caulimoviruses and badnaviruses, the numbering of the CVMV genome was started at the 5' end of this putative primer-binding site (Hull *et al.*, 1986; Medberry *et al.*, 1990).

The tRNA-binding site was followed by an open

reading frame (designated ORF I) that potentially encoded a protein that consisted of 1372 amino acids and had a molecular mass of 186 kDa. The putative protein encoded by ORF I shared only limited similarity with the proteins encoded by the ORF II, III and IV of cauliflower mosaic virus (CaMV). The predicted proteins encoded by ORF IV of the caulimoviruses and the ORF III of the

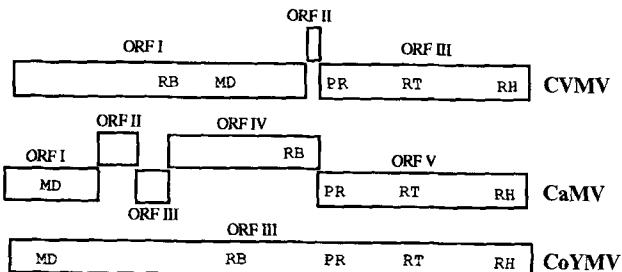


Fig. 3. Comparisons between the genomic organization and relative position of regions of biological importance in CVMV, CaMV and CoYMV. The regions are the intercellular transport movement domain (MD), the RNA-binding domain (RB), the protease motif (PR), the reverse transcriptase motif (RT) and the RNase H activity domain (RH).

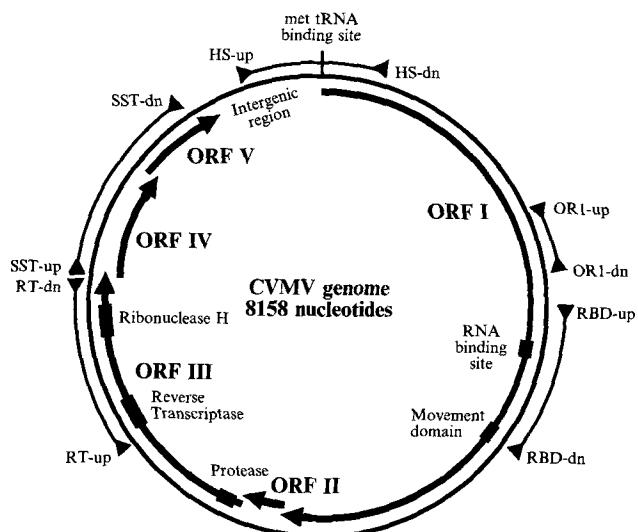


Fig. 4. Predicted genomic organization of CVMV based upon the nucleotide sequence of cloned viral DNA. The circle represents the CVMV genome and the thick black lines with arrows represent the ORFs. The regions of the predicted proteins that contain conserved motifs are represented as blocks. The thin black lines bordered by triangles represent the primer and products that were amplified from CVMV-infected plants. These are illustrated to show the areas of the genome that have been confirmed from at least two independent sources of CVMV.

badnaviruses contain consensus regions for zinc finger-like RNA-binding domains (Covey, 1986). A putative RNA-binding domain was found in ORF I of CVMV at the location aa 739–754 (Fig. 2). The sequence of CVMV ORF I from aa 520–754 has 33% identical and 20% similar amino acids as compared with CaMV ORF IV. The CaMV ORF IV encodes the coat protein and it is probable that the coat protein of CVMV is part of a polyprotein encoded by ORF I of CVMV. The zinc finger-like RNA-binding domain was followed by a region that has similarity with the intercellular transport protein of the caulimoviruses (Figs 3 and 4). The ORF I region aa 972–1105 has 35% identical and 32% similar

amino acids with the figwort mosaic virus (FMV) ORF I aa 96–206. This region contains motifs (Fig. 2) that are typical of intercellular transport proteins (Koonin *et al.*, 1991).

Starting at nucleotide 4132 overlapping 14 nucleotides of ORF I, there was a possible small ORF (designated ORF II) with the potential to encode a protein of 71 amino acids with a molecular mass of 8.8 kDa (Fig. 4).

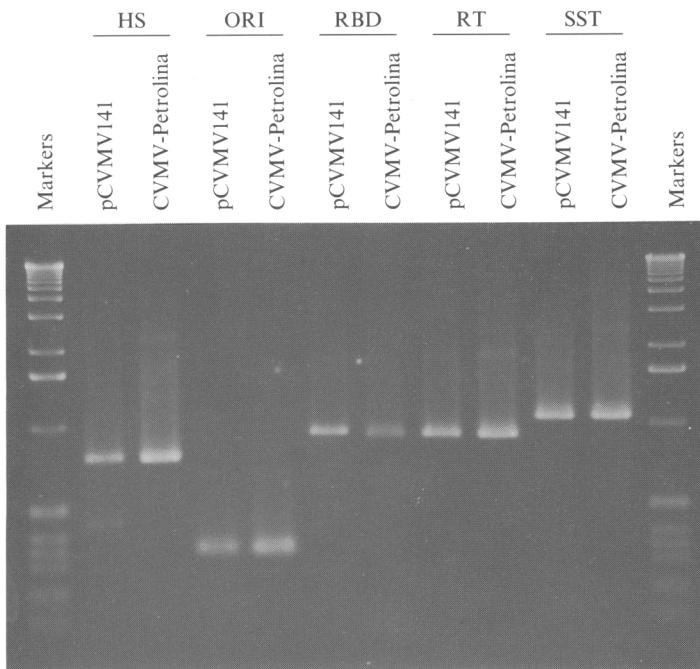
The CVMV ORF III had the potential to encode a protein of 652 amino acids with a molecular mass of 77 kDa. The putative protein contained protease, reverse transcriptase and RNase H consensus regions (Fig. 4). These features made it probable that the ORF III encodes a viral replicase that is typical of the pararetroviruses. The region of the protein that contains the reverse transcriptase motif was compared with the similar proteins of the badnaviruses and caulimoviruses; CVMV was most similar to the caulimoviruses CaMV, FMV and carnation etched ring virus (CERV), with amino acid identities of 47%, 44% and 43%, respectively. When the reverse transcriptase motif of CVMV was compared to those of the badnaviruses, there was a slightly lower level of amino acid identity (39–40%).

The beginning of ORF IV overlapped the end of ORF III and potentially encodes a protein of 201 amino acids with a molecular mass of 24 kDa (Fig. 4). The beginning of ORF V overlapped the end of ORF IV and potentially encodes a protein of 220 amino acids with a molecular mass of 26 kDa. These proteins had no detectable similarity with any of the proteins encoded by the caulimoviruses or badnaviruses.

Following ORF V was a 736 bp intergenic region that contained a putative TATA box (underlined in Fig. 1) which shared similarity with the 3SS promoter of CaMV and the putative promoter of rice tungro bacilliform virus (RTBV).

The most unusual feature of CVMV is the arrangement of genes in the genome (Fig. 3). Based on the sequence data from clone pCVMV141, five sets of primers were used to amplify total nucleic acids from CVMV-infected cassava. In all cases, the amplified products from both the clone pCVMV141 and field isolates of CVMV were of the same size (Fig. 5). A part of the genome of a CVMV isolate from Araripina, Pernambuco, was cloned and the sequence was determined. The Araripina isolate shared 96% nucleic acid and 96% amino acid identity with the pCVMV141 clone. One set of primers (RBD-up and RBD-dn) amplified a product that included the RNA-binding site and the intercellular transport domain. These experiments helped confirm the unusual gene order by confirming that the original clone was very similar to CVMV field isolates.

Comparisons of regions of the proteins that contain the conserved motif for reverse transcriptase were made



Corresponding bases in CVMV genome			Primer Sequence
HS-up	7724 to 7747	5'-GAGTGAGTAGTTCTTAATTCTTC	
HS-dn	349 to 327	5'-CTATCAGCTAAATTCTCTAGC	
OR1-up	1315 to 1334	5'-TAAGAAGAGAAAGGCAATAT	
OR1-dn	1646 to 1627	5'-TGGTGCATTCCCTCTAATT	
RBD-up	2077 to 2094	5'-GCTATTATGAAGTAGGAAAATATGGATG	
RBD-dn	3027 to 3000	5'-TTCTTATTTTATTTGATCCTTGTCAAGG	
RT-up	5343 to 5375	5'-CCATTGGATATCATAACTCACCAAGTATTTC	
RT-dn	6248 to 6216	5'-CAAATAATCAGCAATAATATITGTTTTCCATC	
SST-up	6247 to 6274	5'-TGAGCAGGTACAATTCTGATACTGA	
SST-dn	7332 to 7311	5'-CTTCGTATTCTGGCAGTATAGG	

for CVMV, three badnaviruses and four caulimoviruses. The analysis was made using the computer program PAUP [Phylogenetic analysis using parsimony, version 3.0; developed by D. L. Swofford (Illinois Natural History Survey, Champaign, Ill., USA)]. Heuristic search, branch swapping and majority-rule consensus options of the program were used for making the phylogenetic tree (Felsenstein, 1985). The regions encompassed CVMV ORF III aa 248–427, CaMV ORF V aa 288–468 (Gardner *et al.*, 1981), FMV ORF V aa 280–460 (Richins *et al.*, 1987), CERV ORF V aa 269–451 (Hull *et al.*, 1986), soybean chlorotic mottle virus (SoyCMV) ORF V aa 244–426 (Hasagawa *et al.*, 1989), Commelina yellow mottle virus (CoYMV) ORF III aa 1450–1630 (Medberry *et al.*, 1990), RTBV ORF III aa 1226–1405 (Qu *et al.*, 1991; Hay *et al.*, 1991) and sugarcane bacilliform virus (ScBV) ORF III aa 1302–1482 (Bouhida *et al.*, 1993). Also, the direct amino acid similarity was determined for these regions.

Based upon the amino acid sequences including the consensus sequences of the reverse transcriptase, the

Fig. 5. Comparison of PCR products amplified from the cloned viral DNA (pCVMV141) and from total nucleic acids extracted from CVMV-infected cassava plant extracted from fresh tissue in Brazil (CVMV-Petrolina). The numerical position of the primers corresponds to those reported in Fig. 1 (the complete sequence of CVMV). The approximate positions of these primers are shown on the genomic map in Fig. 2. The markers are the 1 kb ladder (BRL).

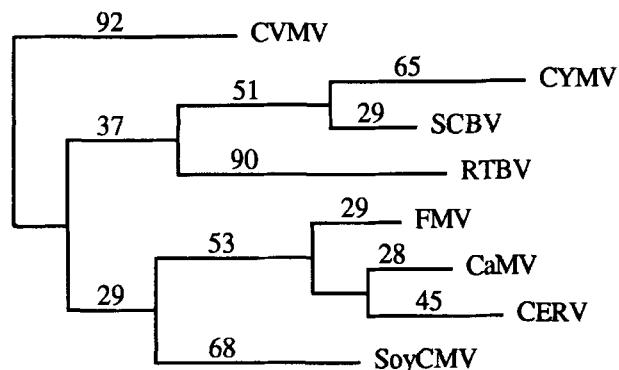


Fig. 6. Diagram illustrating the relationship between a portion of the plant pararetrovirus reverse transcriptases. Amino acid sequences of CVMV, FMV, CaMV, CERV, SoyCMV, RTBV, ScBV and CoYMV were used to construct the tree. The mean assigned branch lengths are written next to each branch.

predicted relationships among eight plant pararetroviruses were determined (Fig. 6). The results of the analysis of the relationship between the plant pararetro-

viruses were similar to those obtained previously (Bouhida *et al.*, 1993). The three badnaviruses and the four caulimoviruses each formed a group. CVMV was consistently placed between the caulimoviruses and badnaviruses. Based on direct amino acid analysis of the intercellular transport motif and the polymerase gene, CVMV appears most closely related to CaMV and FMV.

Since CVMV has consensus sequences for a tRNA^{met}_i-binding site, a zinc finger-like RNA-binding site, an intercellular transport domain, an aspartic acid protease site, a reverse transcriptase motif and an RNase H activity motif, it appears to be a typical plant pararetrovirus. Although CVMV has the morphology of a caulimovirus with an isometric virion of 50 nm, the predicted genomic organization is quite different. The most unusual feature of CVMV is the order of the zinc finger-like RNA-binding motif (probably part of the capsid protein), the putative intercellular transport domain and replicase gene (Fig. 3). In the caulimoviruses, the capsid protein (containing a RNA-binding motif) is encoded by ORF IV, and the replicase protein is encoded by ORF V (Hull *et al.*, 1987). In the badnaviruses, the putative product of ORF III is apparently a polyprotein that is processed into the capsid protein that contains a zinc finger-like RNA-binding sequence, followed by the consensus motifs for a viral aspartic acid protease, the reverse transcriptase and RNase H (Medberry *et al.*, 1990; Qu *et al.*, 1991; Bouhida *et al.*, 1993). In the caulimoviruses and badnaviruses, these genes for these proteins may be translated and the products processed in different manners, but they maintain their relative positions in the genome. In contrast, the CVMV ORF I contains the consensus sequence for the zinc finger-like RNA-binding domain followed by a consensus motif for the putative intercellular transport protein. The sequences for the putative coat protein and intercellular transport protein, which are apparently translated as a polyprotein, precede the replicase gene on the genome. This genomic organization differs from any other plant pararetrovirus that has been reported to date.

We conclude that CVMV is widespread throughout the semi-arid region of north-east Brazil, and based on the predicted genomic organization that CVMV is a plant pararetrovirus that is an atypical caulimovirus.

The authors are indebted to Okon Ansa who prepared the original clones of CVMV in 1981 and Il-Ryong Choi who transferred the viral genome to the cloning vector designated pJAW141 (pUC119 modified to contain a unique *Bgl*II site). The authors are also indebted to M. T. Lin and E. E. Kitajima of the University of Brasilia who supplied CVMV-infected stem cuttings of cassava as a source of virus for cloning. The authors are also indebted to C. Fukuda of EMBRAPA/CPNMF who often accompanied the surveys for CVMV and supplied virus-infected material for the comparison with the original cDNA clone. This work was partially funded by a grant from United Nations Development Projects.

References

- ANZA, O. A. (1982). *Replication and mapping of caulimoviruses*. PhD thesis, University of California-Davis, USA.
- BOUHIDA, M., LOCKHART, B. E. L. & OLSZEWSKI, N. E. (1993). An analysis of the complete sequence of a sugarcane bacilliform virus genome infectious to banana and rice. *Journal of General Virology* **74**, 15–22.
- COVEY, S. N. (1986). Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein of cauliflower mosaic virus. *Nucleic Acids Research* **14**, 623–633.
- DELLAPORTA, S. L., WOOD, J. & HICKS, J. B. (1983). A plant DNA minipreparation: version II. *Plant Molecular Biology Reporter* **1**, 19–21.
- FELSENSTEIN, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- GARDNER, R. C. & SHEPHERD, R. J. (1980). A procedure for rapid isolation and analysis of cauliflower mosaic virus DNA. *Virology* **106**, 159–161.
- GARDNER, R. C., HOWARTH, A. J., HAHN, P., BROWN-LUEDI, M., SHEPHERD, R. J. & MESSING, J. (1981). The complete nucleotide sequence of an infectious clone of cauliflower mosaic virus by M13mp7 shotgun sequencing. *Nucleic Acids Research* **9**, 2871–2888.
- GUILFOYLE, T. J. (1987). Retrolike viruses in plants. In *PlantMicrobe Interactions: Molecular and Genetic Perspectives*, pp. 327–358. Edited by T. Kosuge & E. W. Nester. New York: Macmillan Publishing.
- HASAGAWA, A., VERVER, J., SHIMADA, A., SAITO, M., GOLDBACH, R., VAN KAMMEN, A., MIKI, K., KAMEYA-IWAKI, M. & HIBI, T. (1989). The complete sequence of soybean chlorotic mottle virus DNA and the identification of a novel promoter. *Nucleic Acids Research* **17**, 9993–10012.
- HAY, J. M., JONES, M. C., BLAKEBROUGH, M. L., DASGUPTA, I., DAVIES, J. W. & HULL, R. (1991). An analysis of the sequence of an infectious clone of rice tungro bacilliform virus, a plant pararetrovirus. *Nucleic Acids Research* **19**, 2615–2621.
- HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.
- HULL, R. (1984). Caulimovirus group. *CMI/ABB Descriptions of Plant Viruses*, no. 299.
- HULL, R., SADLER, J. & LONGSTAFF, M. (1986). The sequence of carnation etched ring virus DNA: comparison with cauliflower mosaic virus and retroviruses. *EMBO Journal* **5**, 3083–3090.
- HULL, R., COVEY, S. N. & MAULE, A. J. (1987). Structure and replication of caulimovirus genomes. *Journal of Cell Science (Suppl.)* **7**, 213–229.
- KITAJIMA, E. W. & COSTA, A. S. (1966). Partículas esferoidais associadas ao vírus do mosaico das nervuras da mandioca. *Bragantia* **25**, 211–221.
- KOONIN, E. V., MUSHEGIAN, A. R., RYABOV, E. V. & DOLJA, V. V. (1991). Diverse groups of plant RNA and DNA viruses share related movement proteins that may possess chaperone-like activity. *Journal of General Virology* **72**, 2895–2903.
- LIN, M. T. & KITAJIMA, E. W. (1980). Purificação e serologia do vírus do mosaico das nervuras da mandioca. *Fitopatologia Brasilera* **5**, 419 (abstract).
- LOCKHART, B. E. L. (1990). Evidence for a double-stranded circular DNA genome in a second group of plant viruses. *Phytopathology* **80**, 127–131.
- MASON, W. S., TAYLOR, J. M. & HULL, R. (1987). Retroid virus genome replication. *Advances in Virus Research* **32**, 35–96.
- MEDBERRY, S. L., LOCKHART, B. E. L. & OLSZEWSKI, N. E. (1990). Properties of Commelina yellow mottle virus's complete DNA sequence, genomic discontinuities and transcript suggest that it is a pararetrovirus. *Nucleic Acids Research* **18**, 5505–5513.
- QU, R., BHATTACHARYYA, M., LACO, G. S., DE KOCHKO, A., RAO, B. L. S., KANIEWSKA, M. B., ELMER, J. S., ROCHESTER, D. E., SMITH, C. E. & BEACHY, R. N. (1991). Characterization of the genome of rice tungro bacilliform virus: comparison with Commelina yellow mottle virus and caulimoviruses. *Virology* **185**, 354–364.
- RICHINS, R. D., SCHOLTHOF, H. B. & SHEPHERD, R. J. (1987). Sequence of figwort mosaic virus DNA (caulimovirus group). *Nucleic Acids Research* **15**, 8451–8466.

- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* **74**, 5463–5467.
- SPRINZL, M., HARTMAN, T., MEISSNER, F., MOLL, J. & VORDERWULBECKE, T. (1987). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Research (Suppl.)* **15**, r53–r188.
- TEMIN, H. M. (1989). Retrons in bacteria. *Nature* **339**, 254–255.
- TOH, H., HAYASHIDA, H. & MIYATA, T. (1983). Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis virus and cauliflower mosaic virus. *Nature* **305**, 827–829.
- VIEIRA, J. & MESSING, J. (1987). Production of single-stranded plasmid DNA. *Methods in Enzymology* **153**, 3–11.

(Received 26 October 1994; Accepted 5 January 1995)