Arch Virol (1998) 143: 1723–1744

Archives of Virology

Beak and feather disease virus and porcine circovirus genomes: intermediates between the geminiviruses and plant circoviruses

F. D. Niagro[1], **A. N. Forsthoefel**[2], **R. P. Lawther**[2], **L. Kamalanathan**[1], **B. W. Ritchie**[1], **K. S. Latimer**[1], and **P. D. Lukert**[1]

[1]Psittacine Disease Research Group, College of Veterinary Medicine, University of Georgia, Athens, Georgia, U.S.A.
[2]Department of Biological Sciences, University of South Carolina, Columbia, South Carolina, U.S.A.

Accepted April 4, 1998

Summary. Circoviruses are a diverse group of animal and plant pathogens with undefined relationships to one another but for their non-geminate, non-enveloped capsids and circular, single-stranded DNA genomes. The sequences of the beak and feather disease virus and porcine circovirus genomic DNAs are presented and analyzed in the context of the other members of the family. Sequence comparisons, inferred phylogenies, and geographic occurrence suggest that the ambisense circoviruses, particularly the beak and feather disease virus, represent an evolutionary link between the geminiviruses and the plant circoviruses. We propose that the family members be reclassified into three groups: The family *Circoviridae* consists of the animal pathogens (beak and feather disease virus and porcine circovirus) that possess ambisense genomes with striking similarities to the geminiviruses. The BBTV-like viruses include the plant pathogens (coconut foliar decay virus, banana bunchy top virus, subterranean clover stunt virus) with a geminivirus-like stem-loop element in their DNAs, and single to multiple component genomes. The chicken anemia virus is an unassigned virus possessing unique characteristics bearing little similarity to the other ssDNA viruses.

Introduction

The Sixth Report of the International Committee for the Taxonomy of Viruses identified six viruses as members of the *Circoviridae* [35]. The animal viruses included chicken anemia virus (CAV; the type species), beak and feather disease virus (BFDV), and porcine circovirus (PCV). Three plant pathogens included as unclassified members of the family were the coconut foliar decay virus (CFDV), banana bunchy top virus (BBTV), and subterranean clover stunt virus (SCSV).

Table 1. Characteristics of circoviruses

Virus	Host	Vector	Virion size (nm)	Virion associated proteins
CAV [60]	chicken	Na	23.5	50 kDa capsid/Rep (VP1)
BFDV [60]	psittaciformes	NA	17	26 kDa (VP 1) 23 kDa (VP 2) 16 kDa (VP 3)
PCV [60]	swine	NA	17	36 kDa
CFDV [52]	coconut	planthopper	20	not reported
BBTV [17]	banana	aphid	18–22	21 kDa coat
SCSV [12]	clover, broad bean	aphid	17–19	19 kDa coat

These plant viruses have been described as having properties similar to the geminiviruses. Recent discussions have suggested that one or more of them should be grouped in the "nanavirus" genus [5, 10, 55, 64].

The members of the *Circoviridae* exist as round, non-enveloped virions with mean diameters from 17 to 23.5 nm (Table 1). All members have restricted host ranges. The plant pathogens, like the geminiviruses, require specific insect vectors for transmission. All circoviruses possess circular, single-stranded DNAs that represent the smallest viral DNA replicons known (Table 2). Two of the plant circoviruses possess multipartite genomes, with most components required for infectivity.

DNA sequences have previously been published for the plant circoviruses and CAV. A partial sequence for BFDV has previously been determined and used in the development of DNA based diagnostic tests [16, 23, 33, 51]. In the course of determining the BFDV genomic sequence, the PCV genomic sequence was also completed in this laboratory. Both sequences are presented in this report, although the sequence of the PCV DNA has also recently been reported by two other laboratories [9, 40]. In this paper we report the nucleotide sequence of the beak and feather disease virus and a porcine circovirus isolate with an analysis of the genome organizations and sequence similarities. A discussion of the possible phylogenetic relationships among the circoviruses and geminiviruses is presented to review the family members and to suggest a refined taxonomy. We propose that the BFDV and PCV are members of the *Circoviridae* (with PCV as the type species) and that the chicken anemia virus and the plant circoviruses belong to new separate families.

Materials and methods

BFDV DNA sequence

Animals used in this study were humanely euthanized, diseased birds submitted to the University of Georgia College of Veterinary Medicine for psittacine disease research. Virus was purified from the skin of diseased birds according to published procedures [53]. Virus suspensions were extracted sequentially with equal volumes of Tris-saturated phenol and chloroform-isoamyl alcohol (24:1), and precipitated with ethanol. Viral DNA was digested

Table 2. Characteristics of the circovirus genomes

Virus	Genome component (Accession no.)	Size (bp)	Organization[a]	Putative gene products (ORF in Fig. 2)
CAV [45]	1 (M55918)	2319	+strand	51.7 kDa Rep/capsid (V1) 24.1 kDa (V2) 13.3 kDa (V3)
CFDV [55]	1 (M29963)	1291	+strand	33.5 kDa Rep (V1) 17.2 kDa (V2) 6.7 kDa (V3) 6.4 kDa (V)
BBTV [10]	1 (S56276)	1111	+strand	33.6 kDa Rep (V1)
	2 (L41576)	1060	+strand	none
	3 (L41574)	1075	+strand	20.1 kDa coat protein (V1)
	4 (L41575)	1043	+strand	13.7 kDa (V1)
	5 (L41578)	1018	+strand	19 kDa (V1)
	6 (L41577)	1089	+strand	17.4 kDa (V1)
SCSV [5]	1 (U16730)	1001	+strand	12.7 kDa (V1)
	2 (U16731)	1022	+strand	32.5 kDa Rep (V1)
	3 (U16732)	991	+strand	19.1 kDa (V1)
	4 (U16733)	1002	+strand	17.7 kDa (V1)
	5 (U16734)	998	+strand	18.7 kDa coat protein (V1)
	6 (U16735)	1017	+strand	33.5 kDa Rep (V1)
	7 (U16736)	988	+strand	16.9 kDa (V1)
PCV	1 (AF071879)	1758	ambisense	35.7 kDa Rep (V1) 27.5 kDa capsid (C1) 23.2. kDa (C2) 9.8 kDa (V2)
BFDV	1 (AF071878)	1993	ambisense	34.5 kDa Rep (V1) 28.7 kDa capsid (C1) 17 kDa (V2)

[a] Putative coding capacity of the genome, with respect to the encapsidated strand. Putative gene products are estimated from the inferred amino acid sequences of major ORFs (25 or more amino acids)

with restriction endonuclease Hha-I. The digestion products were separated by electrophoresis. Bands were excised from the gel and the DNA fragments extracted using a crush and soak method. Individual eluted DNA fragments were end labeled and sequenced by the chemical cleavage method [2]. The sequence from one Hha-I fragment was used to design a pair of oppositely oriented primers for amplification of the remainder of the viral DNA by inverse PCR. Primers FN1 (5′-GCTCTAGATGATCCTCCGGAGATTACG-3′) AND FN2 (5′-GCTCTAGAGGGAACCATGCCGTCCAA-3′) were designed with restriction endonuclease Xba-I cleavage sites at their 5′ ends, for subsequent cloning of the amplimer into pUC-19. The full sequence was determined by a deletional subcloning and by sequencing of PCR products from several virus isolates. Sequence runs were assembled into the consensus

sequence using manual methods and the fragment assembly programs in the GCG suite of programs [19].

PCV DNA sequence

The PCV-infected porcine cell line PK-15 (ATCC CCL-33) was obtained from the American Type Culture Collection. Cells were grown in Earle's minimal essential medium with 10% fetal bovine serum at 37 °C in 5% CO_2. DNA from cultured cells and tissues was purified using two rounds of proteinase K digestion and salt precipitation. PCR was performed to amplify viral DNA which might be similar to the BFDV genome. Primers were designed based on amino acid sequence conservation among the reported plant circovirus and the beak and feather disease virus sequences. Using primers FN118 (5′-CCCCTCTTTACTGCAGTA-3′) and FN119 (5′-TGTCGCCGTTGGTGTTTC-3′), designed from the reverse-translation of conserved amino acid sequences, an arbitary primer thermal cycling profile was used to amplify PCV DNA corresponding to the potentially conserved region of sequence. A 900 base pair fragment was labeled with digoxigenin and used to probe unamplified DNA isolated from PK-15 cells and from the uninfected control ESK-4 cells (American Type Culture Collection, Rockville, MD, USA, CL 184). The probe was also used in an in situ hybridization experiment on PCV infected PK-15 cells and a line of PK-15 cells developed at the National Animal Disease Laboratory, Ames, IA, USA, which were identified as PCV negative. The results of that experiment suggested that a small percentage of PK-15 infected (ATCC) cells contained nucleic acid complementary to the 900 base pair probe. This corresponded to the small percentage of cells previously demonstrated to exhibit PCV antigen, using immunocytochemical techniques [26]. The 900 base pair fragment was sequenced and several oligonucleotides were synthesized for use as PCR primers. Amplified DNA fragments were sequenced by dye terminator cycle sequencing. Complementary sequencing runs from PCR generated PCR fragments were aligned using the ALIGN program. Sequence runs were manually assembled into the complete contiguous sequence. PCR primers were chosen with the help of the GENE RUNNER program (Hastings Software, Hastings, NY).

Sequence analyses

Published sequences for circoviruses were obtained from the GenBank sequence database through the WWW server at the NCBI (http://www.ncbi.nlm.nih.gov). Sequence alignments were performed using either the ALIGN program [42] or the PILEUP program in the GCG suite [19]. Sequence similarity searches were performed using the BLAST e-mail server at the NCBI (blast@ncbi.nlm.nih.gov [1]). Inferred phylogenies were constructed using programs in the Phylogeny Inference Package [21].

Genome maps and sequence comparisons

The analysis of the genomes for the circoviruses is based on a descriptive analysis of the published sequences. For comparative purposes, the numbering systems have been changed from those previously published for the viral strand nucleic acid sequences. Numbering begins with the first nucleotide base in the conserved loop sequence for all viral DNA strands. In the convention used for the geminiviruses [7], the encapsidated DNA strand is referred to as the viral strand, encoded ORFs and inferred gene product protein sequences are indentified as V1, V2, etc. Conversely, the complementary (unencapsidated) DNA strand ORFs as C1, C2, etc.

a

```
   1  TTAGTATTAC CCGCCGCCTG GGGCACCGGG GCACCGCAGC TATTGGCTGC CTTGCGGAGG TGCCCGCCCC TAGGGAGGAG TAAATGGCGC CGTTAGGCGC
 101  CCGTTAATCT CCGGAGGATC ACACTCGTCC GGGAACCATC CCGTCCAAGG AGGGCTCTGG CTGTCGCCGT TGGTGTTTCA CTCTTAACAA CCCTACAGAC
 201  GGCGAAATCG AATTCGTCCG TTCTCTCGGG CCTGACGAAT TCTACTATGC CATCGTTGGA CGGGAAAAGG GTGAGCAAGG CACTCCCCAT TTGCAAGGCT
 301  ACTTTCATTT TAAAAATAAG AAGCGACTGA GCGCGCTTAA GAAAATGCTG CCGCGAGGTC ATTTTGAGCG CGCTAAAGGG AGCGACGCGG ATAACGAGAA
 401  GTATTGCAGT AAAGAGGGGG ACGTCATACT TACCCTGGGC ATTGTGGCGA GAGACGGTCA CCGCGCTTTC GATGGAGCTG TTGCTGCCGT GATGTCCGGA
 501  CCCAAAATGA AGGAAGTCGC GCGAGAGTTC CCAGATATCT ACGTCAGGCA TGGGCGGGGC TTGCATAGCC TCTCGCTATT GGTTGGTTCC CGCCCGCGTG
 601  ATTTTAAGAC TGAGGTTGAC GTCATTTACG GGCCACCGGG GTGTGGCAAG AGTCGATGGG CCAATGAGCA GCCTGGGACC AAATATTATA AAATGCGCGG
 701  TGAATGGTGG GATGGATATG ATGGGGAAGA TGTCGTCATC TTGGACGACT TTTATGGGTG GCTACCTTAT TGCGAGATGC TCCGCCTCTG CGACCGTTAC
 801  CCACATAAAG TGCCAGTTAA GGGCGCTTTT GTGGAGTTTA CGAGCAAGAG GATCATTATC ACGAGCAATA AGGCCCCGGA GACCTGGTAC AAGGAGGACT
 901  GTGACCCGAA GCCACTGTTC CGGAGATTCA CTCGTGTTTG GTGGTACAAC ATTGACAAGT TGGAACAAGT CCGGCCTGAT TTCCTCGCCC ACCCCATCAA
1001  TTTTTGATAG TCTCGGTGAT GTTTAATAAA GTTAGGCGTC GGGCCGAAGG CCCGATGCCG CAGGGGGGGA CCCCCTGCCG GAGGGTTCGC AGGGCCGTCA
1101  GGCCCGAGAA CCCGACCAGC CCGGAGGGCC TAGTCTGTAT CGGGGGGGGG GCCCCGGGGG GTCCCCCCGA CACGACGAAC ACGGTAGCGA CGAAGGCGCC
1201  AATAAACACT CAAAAAGGTA TTTGCTGCTT GAGTCTTTAT TAAGTACTGG GATTGTTAGG GGCAAACTGA CGGAATTGAA CATATATAGT GAGCTTGGTT
1301  ACATAAGTGA TCGTTTGTTC TGGCTGAGGG AAGCTGAAGC CAATGCCGTA GTGCCTGACT TTCGCTCCTG CTGCGTTGGG TCCTCCTTGT AGTGGGATCC
1401  AGCCGGTTCT GGCGCTGTTT AGCCACAATG CTGCAGACTG GTTCGCTGTG GTGAGGTCGT TTATCGTTAT TTGTGGTTTT GGTCTGAGGA GACGTTTGAA
1501  TCCCCTGCTA ACATACCATT TTCTGGCACC GTCGAAAGGT GCCAGTGGGT CTTGTGTTTG GTCTGCAACA GTTTTAAATT TACTTATCCT GGAGTCTTGG
1601  ATTACGGCCG TGTGGCCGAA TCCGTCTCCT TGTATGGTGT AGTGTCCCCA TGTGGGCCTC ATTTCCATTT TAGCTAACTT AATCCGGTAG TTTTCGAAAT
1701  TCAGTGTTTG TGGGTTTGGT GTGTTTGTTA TGAAGTCTGA CAATGCAAAG GTTACAAAGT CGGAGCTAAA AATTAGGTTG CCAGTACTGG TGGTTTGTTT
1801  TTGAATTTGG AATTTGAATT GGCGTGTGAG TCTGAGAGTG TAAACCCTAT TGGTTGAGAA ACGGCGTCTG CGGAAGTGTC TACGTCGCCG GCGGTATCGC
1901  CTGATGTGAC GTCTGCGGTA GTATGGGCGG GCATACCGTC GTCTAACCTG AAATATAGCG CATGCGTAGT TAGAGGTGCC CCATAGGCGG CGG
```

b

```
   1  GTAGTATTAC CAGCGCACTT CGGCAGCGGC AGCACCTCGG CAGCGTCAGT GAAAATGCCA AGCAAGAAAA GCGGCCCGCA ACCCCATAAG AGGTGGGTGT
 101  TCACCCTTAA TAATCCTTCC GAGGAGGAGA AAAACAAAAT ACGGGAGCTT CCAATCTCCC TTTTTGATTA TTTTGTTTGC GGAGAGGAAG GTTTGGAAGA
 201  GGGTAGAACT CCTCACCTCC AGGGGTTTGC GAATTTTGCT AAGAAGCAGA CTTTTAACAA GGTGAAGTGG TATTTTGGTG CCCGCTGCCA CATCGAGAAA
 301  GCGAAAGGAA CCGACCAGCA GAATAAAGAA TACTGCAGTA AAGAAGGTCA CATACTTATC GAGTGTGGAG CTCCGCGGAA CCAGGGGAAG CGCAGCGACC
 401  TGTCTACTGC TGTGAGTACC CTTTTGGAGA CGGGGTCTTT GGTGACTGTA GCCGAGCAGT TCCCTGTAAC GTATGTGAGA AATTTCCGCG GGCTGGCTGA
 501  ACTTTTGAAA GTGAGCGGGA AGATGCAGCA GCGTGATTGG AAGACAGCTG TACACGTCAT AGTGGGCCCG CCCGGTTGTG GGAAGAGCCA GTGGGCCCGT
 601  AATTTTACTG AGCCTAGCGA CACCTACTGG AAGCCTAGTA GAAATAAGTG GTGGGATGGA TATCATGGAG AAGAAGTTGT TGTTTTGGAT GATTTTTATG
 701  GCTGGTTACC TTGGGATGAT CTACTGAGAC TGTGTGACCG GTATCCATTG ACTGTAGAGA CTAAAGGCGG TACTGTTCCT TTTTTGGCCC GCAGTATTTT
 801  GATTACCAGC AATCAGGCCC CCCAGGAATG GTACTCCTCA ACTGCTGTCC CAGCTGTAGA AGCTCTCTAT CGGAGGATTA CTACTTTGCA ATTTTGGAAG
 901  ACTGCTGGAG AACAATCCAC GGAGGTACCC GAAGGCCGAT TTGAAGCAGT GGACCCACCC TGTGCCCTTT TCCCATATAA AATAAATTAC TGAGTCTTTT
1001  TTGTTATCAC ATCGTAATGG TTTTTATTTT TATTTATTTA GAGGTCTTTT AGGATAAATT CTCTGAATTG TACATAAATA GTCAGCCTTA CCACATAATT
1101  TTGGGCTGTG GCTGCATTTT GGAGCGCATA GCCGAGGCCT GTGTGCTCGA CATTGGTGTG GGTATTTAAA TGGAGCCACA GCTGGTTTCT TTTATTATTT
1201  GGGTGGAACC AATCAATTGT TTGGTCCAGC TCAGGTTTGG GGGTGAAGTA CCTGGAGTGG TAGGTAAAGG GCTGCCTTAT GGTGTGGCGG GAGGAGTAGT
1301  TAATATAGGG GTCATAGGCC AAGTTGGTGG AGGGGGTTAC AAAGTTGGCA TCCAAGATAA CAACAGTGGA CCCAACACCT CTTTGATTAG AGGTGATGGG
1401  GTCTCTGGGG TAAAATTCAT ATTTAGCCTT TCTAATACGG TAGTATTGGA AAGGTAGGGG TAGGGGGTTG GTGCCGCCTG AGGGGGGGAG GAACTGGCCG
1501  ATGTTGAATT TGAGGTAGTT AACATTCCAA GATGGCTGCG AGTATCCTCC TTTTATGGTG AGTACCAATT CTGTAGAAAG GCGGGAATTG AAGATACCCG
1601  TCTTTCGGCG CCATCTGTAA CGGTTTCTGA AGGCGGGGTG TGCCAAATAT GGTCTTCTCC GGAGGATGTT TCCAAGATGG CTGCGGGGGC GGGTCCTTCT
1701  TCTGCGGTAA CGCCTCCTTG GCCACGTCAT CCTATAAAAG TGAAAGAAGT GCGCTGCT
```

Fig. 1. Complete nucleotide sequence of BFDV and PCV virus strand DNAs. Numbering begins with the first base in the TTAGTATTA loop sequence. The loop sequence in the putative replication origin (underlined and bold) and complementary strand TATA sequence (bold) are highlighted. **a** BFDV sequence contains amino terminal TCT codons (underlined) for both the V1 and C1 ORFs. **b** PCV sequence contains ATG codons (underlined) for both the V1 and C1 ORFs. Sequence lacks the third nucleotide of a GGG triplet at base number 1045, as compared with sequences reported previously [9, 40]

Results

BFDV and PCV sequences

The sequences of the virus-encapsidated DNA strands for BFDV and PCV are presented in Fig. 1. The BFDV and PCV sequences consist of 1993 and 1758 bases, respectively. This PCV sequence differs slightly from those previously reported [9, 40].

The genome of the porcine circovirus appears to possess an ambisense organization (Table 2; Fig. 2a). This analysis is supported by other recently published studies [37, 40]. The viral strand encodes a putative Rep protein (ORF V1) with all

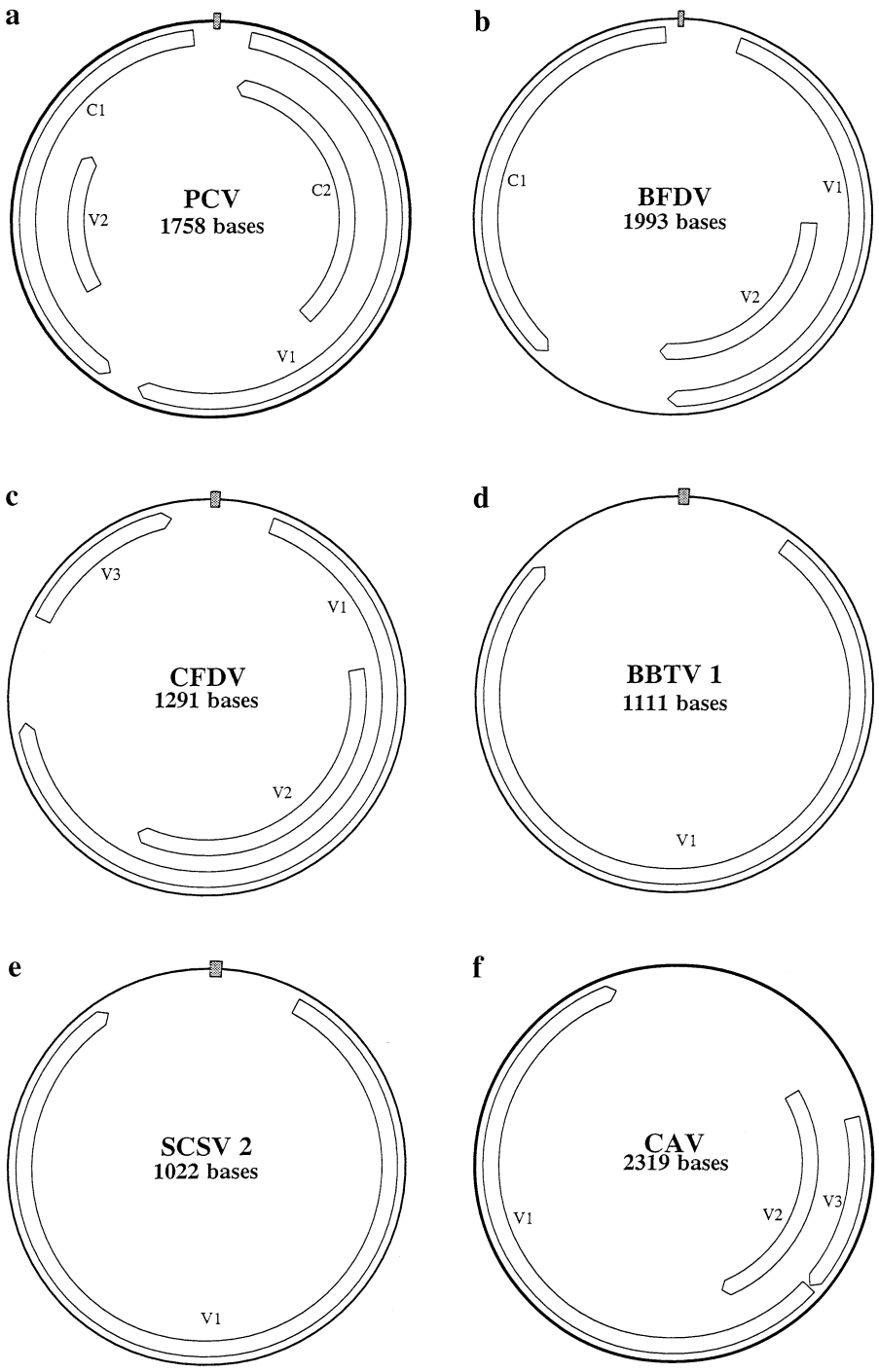


Fig. 2 (*continued*)

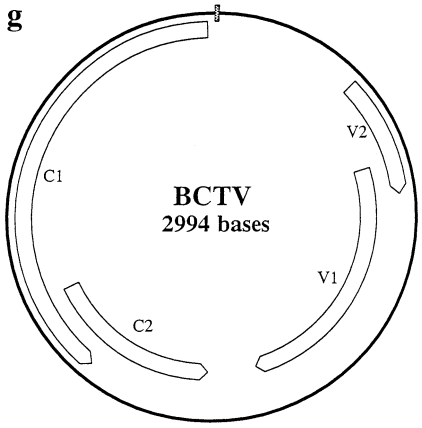


Fig. 2. Physical maps of the circoviruses and a representative geminivirus. Accessions numbers are indicated in Table 2. Maps represent the reported viral (encapsidated) DNA strand. Nucleotides may be renumbered for this comparison (see Materials and methods) and base number one is at the twelve o'clock position. ORFs on the viral strand are labeled as V1, V2, etc.; those on the inferred complementary strand are labeled as C1, C2, etc. The stem-loop structures (or structurally conserved elements) are represented by the box at the top of each circle. **a** Porcine circovirus (PCV). **b** Beak and feather disease virus (BFDV). **c** Coconut foliar decay virus (CFDV). **d** Component 1 of banana bunchy top virus (BBTV). **e** Component 2 of subterranean clover stunt virus (SCSV). **f** Chicken anemia virus (CAV). **g** Beet curly top virus (BCTV), a geminivirus

three sequence motifs found in molecules involved in rolling circle DNA replication and it possesses the P-loop motif (Fig. 3). The inferred complementary strand encodes a putative viral capsid protein from ORF C1. This ORF is flanked by eukaryotic transcription control and polyadenylation signals. The inferred sequence of this putative capsid protein possesses a high percentage of arginine and lysine residues near the amino terminus, and has a high degree of similarity to the protamine block sequence (Fig. 4a). The non-coding region separating the beginnings of both oppositely oriented ORFs contains features suggestive of replication and transcription control functions. Two additional major ORFs are present (one on each strand). Neither is flanked by the appropriate signals for expression, however, evidence for RNA splicing has been reported [36].

The beak and feather disease virus (BFDV) DNA also has an ambisense organization (Table 2; Fig. 2b). Northern blot analyses, using strand-specific oligonucleotides to probe polyadenylated RNA from BFDV infected feather pulp, support this conclusion [44]. The viral strand possesses a major ORF which encodes a putative Rep protein, possessing all rolling circle replication sequence motifs and the P-loop motif (Fig. 3). The inferred amino acid sequence also contains a potential pyrophosphatase domain (Fig. 8 [6]). A second major ORF (V2) is encoded in a different reading frame on the viral strand (Fig. 2b; Table 2). This ORF could encode the 16 kDa virion associated protein (Table 1). The BFDV complementary strand encodes a putative capsid protein with striking similarities to the PCV capsid protein. An alignment of the PCV and BFDV inferred capsid

```
                                                            1                                                  2
                                                           f tl                                             pHuHuuu
consensus                                            R     F-TLN   t  E     r                 Y   G     E     G   pHLQG   fk    k

BFDV      1 SPEDH-----------TRPGTIPSKEGSGCR--RWC--F-TLNNP-TDGEIEFV-R---- 38      47 YAIVG----REKGEQ-G-TPHLQGYFHFK---NKK----- 71
PCV       1 M.SKK-----------SG.--------QPHK--..V--.-......-SEE.KNKI-.---- 30      39 .FVC.----E.GL.E-.R......FAN.A---K.Q----- 65

SCSV      1 --------------------------MA.--.Y.--.-...YA-.EI.R.TF-L---- 21      31 .FV..----D.TATT-.-QK....FVS..---..I----- 56
CFDV      1 ------------------------M..SI.--..--.-...YE-.EE.AAN.V.---- 25      34 ........----D.VAPST.-QR....FI.L.---TGR----- 61
BBTV      1 ------------------------MS.PSL--K..--.-...YS-SAA.R.NF-L---- 24      33 ..V..----D.VAPAT.-QK.....LSL.---KRI----- 60

CSMVAu    1 ---MS-----------SL.VSESEG....TS--VQV--P-SRGGQV.P..KA.S-L---- 36      70 .IYIS----..FHAD-.-E...HAFVQLE---ANF----- 95
ToLCVAu   1 --------------------------MT.--PKS--.-RI.AK-NYFLTYPK-C---- 21      43 .IK.C----..LH.N-.-E...HVLIQ.E---G.FQCKNQ 73
ICMV      1 MSPPKRFQINAKNYFL.Y.RCSLT..EALSQ--IRN--.Q.PT..-KFIK.--------- 45      47 ----C----..LH.N-.-E...HVLIQ.E---G.YKCQNQ 74
SLCV      1 --------------------NM.RNPN.FRLTA.NI--.-LTYPR-C.VPK.E.-LEMLL 35      46 .VR.A----..EHSD-.-S...HCLIQLS---G.SNIKDA 76
TMoV      1 M.--------------------.P.KFRVQS--KNY--.-LTYPQ-CSLSK.EA-L---- 29      43 FIKIC----..LH.N-.-E...HVLVQ.E---G.YQCTNN 73
TYDV      1 M.------------------SA.Q.TK.FRL--QTKYV.-LTYPR-CSSSA.NL-.---- 33      48 I..-A----T.LHQD-.-....HCLIQLDKRS.IR----- 75
AbMV      1 M.-------------------P.K.FRVQAK--NYF--L-.YPQC-SLTKD.AL-S---- 30      38 VNKKFIKIC..LH.N-.-E...HVLIQ.E---G.YQCTNN 74
WDV       1 MASSS-----------APRFRVY..YLFLTY--PE.-----..-E.-QYALDSLR-T---- 36      45 .IAAV----..LH.D-.-S...--HVLVQ---..L----- 69


                                                3                                      P-loop
                                                *
                             ux              x Yuxk                              Gxxxx GKS
Consensus       R H E   akG     sD N             YCsKE                                 G   GKs   A

BFDV       83 RGHFER--AKG---SDADNE-------K-YCSKEGD-- 105          175 ---GPPGC-GKSRWAN---EQPGTK 192
PCV        77 .C.I.K--.....---T.QQ.K-------E-......H--  99          171 ---......-...Q..RNFT.PSD.Y 191

SCSV       68 .A.W.I--.R.----..SQ.R-------D-..C..-T--  89          173 ---SDG.E-..TSF.K---.----- 185
CFDV       74 .I.L.P--TR.----..EQ.R-------D-......-R--  95          185 ---RDG.D-...VF.---------- 195
BBTV       72 .A.W.I--.R.---T.EE.S-------.-......-T--  93          177 ---..Y.NE...TY.K---SLIKKD 195

CSMVAu    106 DEFHPNIQ.AR---QP.STL-------.-..M.HPE-- 130          189 ---FQYSA-NALFPDP---P.TYSA 206
ToLCVAu    84 SA..HP--NIQGAK.SS.VK-------S-.LE.D..-- 109          231 ---.ESRT-..TV..R---SLGPHN 238
ICMV       85 SA..HP--NIQGAK.SS.VK-------S-.ID.D..-- 111          223 ---.DSRT-..TM.SR---ALGPHN 241
SLCV       89 .F.PNI--QAA---K.TNAV-------.N.IT....YC 114          219 ---.GSRT-..TM..R---SLGPHN 236
TmoV       84 SA..HP--NIQGAK.SS.VK-------S-.ID.D..-- 109          219 IVE.DSRT-..TM..R---ALGPHN 239
TYDV       86 HPNIQP--..N---.EQVL.---------.I..D.N-- 107          187 ---HEDLQ-LWRDQHL---YHVSVD 204
AbMV       85 SA..HP--NIQGAK.SS.VK-------S-.ID.D..-- 110          238 ---.----------H.---YLS.HL 249
WDV        77 NALNL.--MDT---.PFSIFHPNIQAA.-D.NQVR.-- 107              -------------------------
```

protein amino acid sequences reveals a highly conserved sequence of 14 amino acids (YVtkLTIYVQFRqF) near the carboxyl terminus of both proteins as well as adjacent corresponding myristylation sites (Fig. 4b). The strict conservation of these two motifs suggests their necessity for protein function.

Discussion

The other viruses in the family

The genome of the coconut foliar decay virus (CFDV) possesses transcriptional start and polyadenylation signals flanking major open reading frames (ORF) on only the virion-packaged DNA strand (Table 2; Fig. 2c). The protein composition of the virion has not been reported (Table 1). However, the major ORF (V1) encodes a putative Rep protein, possessing the three sequence motifs found in rolling circle replication associated proteins (Fig. 3). This putative Rep protein possesses the consensus ATP/GTP binding motif I(P-loop) found in numerous proteins, including those with helicase activity [27]. Open reading frame V4 encodes a putative 6.4 kDa protein with 23% arginine residues (Fig. 4a). Rohde et al. [55] suggested that this is the putative capsid protein, based on sequence similarities with the *Autographa californica* nuclear polyhedrosis virus (AcNPV) core protein and to the amino terminus of several luteovirus coat proteins [61].

Six different DNA components have been found associated with an Australian isolate of the banana bunchy top virus (BBTV [10]). Five of the six components each have one major ORF associated with transcription control signals on the virion-packaged DNA strand (Table 2; Fig. 2d). Component 1 encodes a putative Rep protein that possesses all three consensus motifs and a P-loop sequence (Fig. 3). Both components three and five encode putative proteins of the approximate size of the 21 kDa coat protein (Table 2). However, the sequence encoded by

Fig. 3. Alignment of the inferred amino acid sequences of the putative replication associated proteins (Rep protein) of the circoviruses (BFDV, PCV, SCSV, CFDV, BBTV) with those of several of the geminiviruses. Only the regions which encompass the three Rep protein motifs and the P-loop sequence are shown. Above the aligned sequences are the consensus sequences unique to the circoviruses (Consensus) with the three (1, 2, 3) conserved sequence motifs reported for the initiator and movement proteins involved in rolling circle replication [28]. The putative active site tyrosine (*) is identified. The conserved ATP/GTP binding motif (P-loop [27]) is indicated above the aligned sequences. Viruses and GenBank accession numbers: *BFDV* beak and feather disease virus (AF071878); *PCV* porcine circovirus (AF071879); *SCSV* subterranean clover stunt virus (U16731); *CFDV* coconut foliar decay virus (M29963); *BBTV* banana bunchy top virus (S56276); *CSMVAu* chloris striate mosaic virus Australian isolate (JU0043); *ToLCVAu* tomato leaf curl virus-Australian isolate (P27295); *ICMV* indian cassava mosaic virus (JQ2327); *SLCV* squash leaf curl virus (M38183); *TMoV* tomato mottle virus (M90495); *TYDV* tobacco yellow dwarf virus (M81103); *AbMV* abutilon mosaic virus (A36214); *WDV* wheat dwarf virus (B24356). Dots in aligned sequences represent bases which are identical to those in the aligned BFDV sequence. Numbers indicate amino acid position in sequence. Dashes represent gaps introduced to optimize the alignment. “u” represents bulky hydrophobic residue [28]

the ORF on component two possesses an arginine/lysine-rich region at the amino terminus like the CFDV. Sequences have also been reported for one or more components of BBTV isolates from Taiwan and Hawaii. Sequence analyses suggest that at least two different groups exist: a South Pacific group and an Asian group [17, 29, 63, 64]. The sequence of the Hawaiian isolate is like those in the South Pacific group.

a

```
Japanese Quail        4          YRRRTRTRSRSRRRRRSRRRRSSRRRRY          30
Equine               17           RRRRRRCRRRRRRRSVRQRRVCCRRYTVLRCRRR   49
Rainbow Trout         3  RRASRPVRRRRRPRVSRRRRRGGRRRR                   30
Opossum              21  RRRSRYRSRRRRYRRSRRRRRRGRRRRGYSRRRYSRRGRRR     61
Cuttlefish           15  RRRSRSPYRRRYRRRRRRRRRRSRRRRYRRRRSYSRRRYRR     55

BFDV          SNYACAIFQVRRRYARPYYRRRHIRRYRRRRRHFRRRRFSTNRVYTLRLTRQ-FKFQIQKQTTSTGN LIFSS.....
PCV                     MTWPRRRYRRRRTRPRSHLGNILRRRPYLAHPAFRNRYRWR-RKTGIFNSRLSTEL VLTIK.....
CAV                   MARRARRPRGRFYSFRRGRWHHLKRLRRRYKFRHRRRQRYRRRAFRKAFHNPRPGTYS VRLPN.....
SCSV                          MVAVRWGRKGLRSQRRKYSRIAYKPPSSKVVSHVESVLNKRDVTGAEVK PFADG.....
BBTV                        MFRQEMARYPKKSIKKRRVGRRKYGSKAATSHDYSSSGSIL----VPENTVK VFRIE.....
CFDV                  MRTRRRRREVRVCQISRTQARLVLHMWWNQKGRIVPVHRGPKTKF-----NPRCTQV

BCTV                    MRKYTRNTYTMSQKRKVNPQSAWPKKRRTSTTSRKYQWRRPVTKNRTLKLKMYDDM LGAG
CSMV-Au                                  MSPASSWKRRPSSSSAQASKKRRVYRPAVSRSLARREPL QVQDF.....
ToLCV-Au   MSKRPADIVISTPASKVRRKLNFNSPFKSAAAVPTVRVTRRRTWVNRPMYRKPMMYRLFRSPDVPRGCEG PCKVQ.....
AcNPV 6.9                      MVYRRRRRSSTGTTYGSTRRRRSSGYRRRPGRPRTYRRSRSRSSTGRRSY RTRYY.....

BWYV      MNTVVGRRTINGRRRPRRQTRRPQRNQPVVVVQTSRRTQRRPRRRRRGNNRAGRT.....

SBMV      ATRLTKKQLAQAIQNTLPNPPRRKRRAKRRAAQVPKPTQA........
```

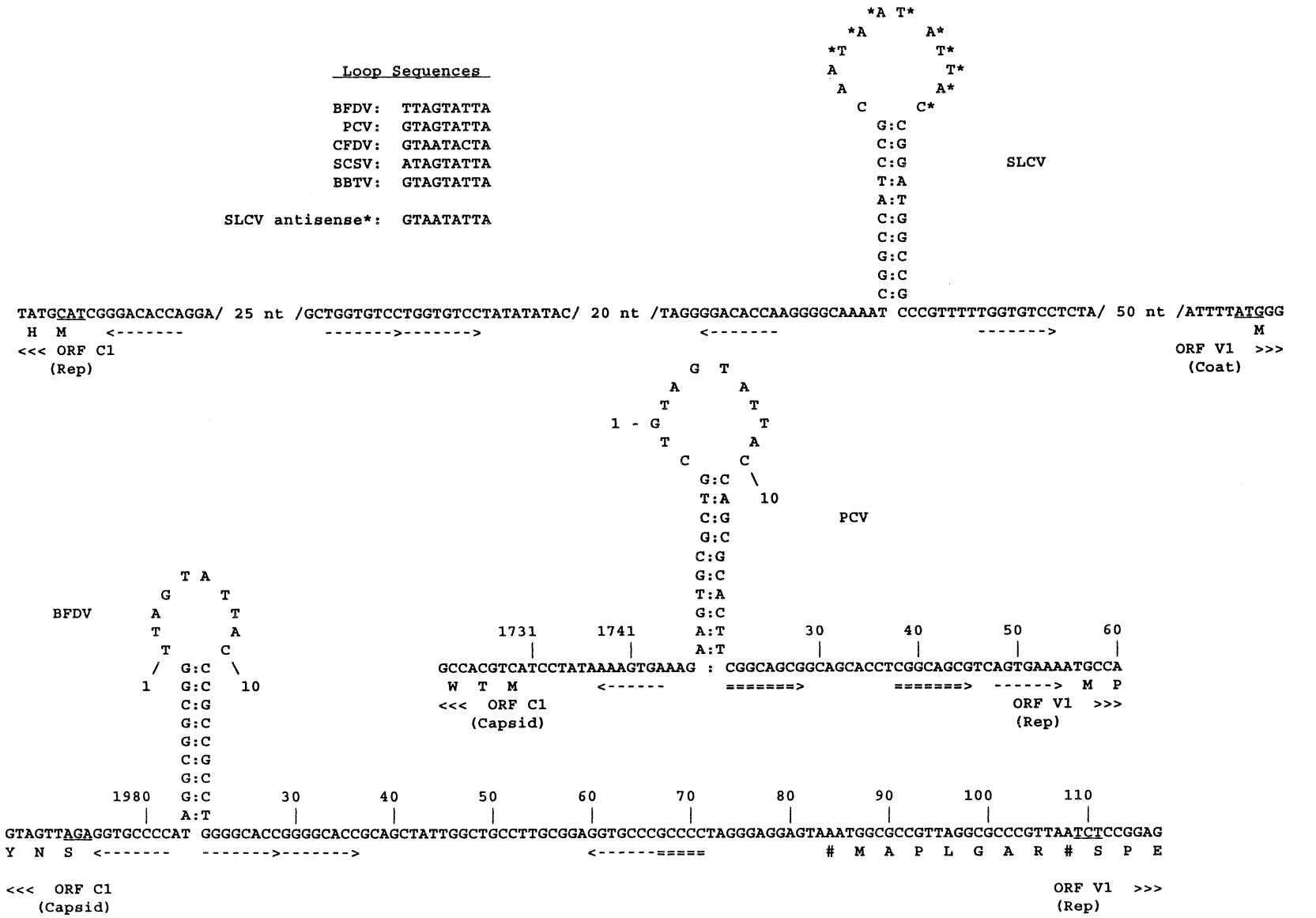
b

```
BFDV        1 SNYACAIFQVRRRYARPYYRRRHIRRYRRRRRHFRRRRFSTNRVYTLRLTRQFKFQIQKQ 60
PCV         1 ------MTWP....R.RRT.P.SHLGNIL...PYLAHPAFR..YRWR.K.GI.NSRLSTE 54

BFDV       61 TTSTGNLIFSSDFVTFALSDFITNTP-------NPQTLNFENYRIKLAKMEMRPTWGHYT 113
PCV        55 LVL.IKGGY.QPSWNVNYLK.NIGQFLPPSGGT..LP.P.QY...RK..Y.FY.-RDPI. 113

BFDV      114 IQGDGFGHTAVIQD-------------------------------SRISKFKTVADQTQD 142
PCV       114 SNQR.V.S.V..L.ANFVTPSTNLAYDPYINYSSRHTIRQPFTYH..YFTP.PEL...I. 173

BFDV      143 PLAPFDGARKWYVSRGFKRLLRPKPQITINDLTTANQSAALWLNSARTGWIPLQGGPNAA 202
PCV       174 WFH.-------------------------------------N.KRNQL.LH.NTHT.-- 193

                    ******          YVtkLTIYVQFRqF
BFDV      203 GAKVRHYGIGFSFPQPEQTITYVTKLTIYVQFRQFAPNNPST 244
PCV       194 ---.E.T.L.YALQNAATAQN..VR........E.--ILKDL 230
```

The legumes of Australia are afflicted by a stunt caused by another multi-component circovirus. Seven genome component DNAs have been sequenced from an isolate of SCSV [5]. Each component of the SCSV encodes a single putative protein from a major ORF on the virion-packaged DNA strand (Table 2; Fig. 2e). Putative Rep proteins are encoded on both components two and six, although the sequence of motif three on the inferred component six protein differs considerably from that of the consensus. Both inferred sequences possess P-loop motifs (Fig. 3). The sequence of the ORF on component five has been identified as the coat protein based on a previous comparison with the amino terminal peptide sequence obtained from purified viral coat protein [13]. The amino terminus of this protein also contains a high percentage of arginine and lysine residues (Fig. 4a).

Unlike the other animal circoviruses, the chicken anemia virus (CAV) does not possess an ambisense DNA. The existence of all three putative major gene products (Table 2) have been confirmed and matched to proteins found in the virion or infected cells through transcription analysis [49]. The proteins are translated from a single, 2.0 kbp polycistronic message with nested open reading frames from only one strand of the replicative form DNA (Fig. 2f). The largest ORF encodes a 52 kDa protein, VP1, which is the capsid protein and which also possesses the 3 amino acid sequence motifs found in the replication associated proteins involved in rolling circle replication. The amino terminal region of that protein possesses an arginine-rich region with significant amino acid identity to the consensus protamine block sequence [39]. The 24 kDa protein (VP2) has been expressed in vitro and is believed to perform a scaffolding function during virion assembly [20]. The 13 kDa protein (VP3), also rich in arginine and lysine residues, has been expressed in vitro and has been demonstrated to induce apoptosis in chicken lymphoblastoid T cells and myeloid cells [46]. Mixtures of in vitro expressed (baculovirus) viral proteins have been used to inoculate chickens

Fig. 4. Alignment of the putative coat protein sequences. **a** Alignment of the amino terminal sequences of the putative coat proteins of the circoviruses (SCSV, BBTV, BFDV, PCV, CAV, and CFDV), three representative geminiviruses (BCTV, CSMV-Au, ToLCV-Au), and the 6.9 kDa core protein of the *Autographa californica* nuclear ployhedrosis virus (AcNPV) to representative protamine block sequences (Japanese quail, P14402; equine, P15341; rainbow trout, P02331; opossum, P35305; and cuttlefish, S14085). Alignment was performed using PILEUP (GCG version 8) for all complete sequences, although only the amino terminal amino acids are shown for most. Dashes represent gaps introduced to optimize the alignment. Basic residues (arginine and lysine) are highlighted. Below the aligned sequences are the amino terminal sequences of a luteovirus coat protein (beet western yellows virus; *BWYV*; X14600) and the sobemovirus, southern bean mosaic virus (*SBMV*, A04218) for which the coat protein structure has been solved. **b** Alignment of the amino acid sequences of the putative coat proteins of the beak and feather disease virus (*BFDV*) and the porcine circovirus (*PCV*). The conserved myristylation sites (G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}; asterisks) and highly conserved 14 amino acid sequence at the carboxyl termini (underlined) are indicated. Numbers indicate amino acid position in sequence


Loop Sequences
BFDV: TTAGTATTA
PCV: GTAGTATTA
CFDV: GTAATACTA
SCSV: ATAGTATTA
BBTV: GTAGTATTA
SLCV antisense*: GTAATATTA
SLCV
PCV
BFDV
TATGCATCGGGACACCAGGA/ 25 nt /GCTGGTGTCCTGGTGTCCTATATATAC/ 20 nt /TAGGGGACACCAAGGGGCAAAAT CCCGTTTTTGGTGTCCTCTA/ 50 nt /ATTTTATGGG
<<< ORF C1 (Rep)
ORF V1 >>> (Coat)
GCCACGTCATCCTATAAAAGTGAAAG : CGGCAGCGGCAGCACCTCGGCAGCGTCAGTGAAAATGCCA
<<< ORF C1 (Capsid)
ORF V1 >>> (Rep)
GTAGTTAGAGGTGCCCCAT GGGGCACCGGGGCACCGCAGCTATTGGCTGCCTTGCGGAGGTGCCCGCCCCTAGGGAGGAGTAAATGGCGCCGTTAGGCGCCCGTTAATCTCCGGAG
Y N S # M A P L G A R # S P E
<<< ORF C1 (Capsid)
ORF V1 >>> (Rep)

for the production of antibody responses. Mixtures containing both VP1 and VP2 or mixtures containing all 3 proteins are required for the elicitation of virus neutralizing antibodies and for protection from challenge by live CAV [32].

Similarities

The non-coding regions on the DNA components of all circoviruses except the CAV display features strikingly similar to those found in the geminiviruses. All geminivirus DNAs possess a so-called structurally conserved element (SCE). This stem-loop structure is involved in both DNA replication and in control of viral gene expression. The most striking similarity is found in a comparison of the PCV and BFDV non-coding regions to those of the geminiviruses (Fig. 5). This comparison reveals that the stem-loop structure (SCE) is flanked by iterated sequences like the so-called iterons described in geminivirus genomes [3]. Another recent analysis has described three copies of a hexanucleotide repeat within the region flanking the stem-loop, with similarity to a geminivirus replication protein binding site [37].

The apparent inverse genome orientation is further supported by a comparison of the BFDV and PCV genome organizations (Fig. 2a and 2b) to that of a representative geminivirus, the beet curly top virus (BCTV), which is shown in Fig. 2g. In the case of the geminiviruses, the major viral strand ORF (V1) encodes the virus coat protein, while the major ORF found on the complementary strand (C1) encodes the replication associated protein (Rep). For the ambisense circoviruses, the opposite is true. The BFDV genome is organized like the PCV genome, except that the putative TATA sequence for the capsid protein gene (ORF C1; Fig. 1a) is located within the ORF, downstream of the putative initiation codon. This striking similarity between the PCV and BFDV non-coding regions (Fig. 5) suggests that a serine codon (TCT) may be used by the psittacine host protein synthesis machinery for initiation of translation, however, this possibility has yet to be confirmed. The presence of serine codons immediately downstream from the distal iterated sequences flanking both sides of the stem-loop region strongly supports this hypothesis. In terms of the organization of the ambisense genome and the non-coding region associated with the stem-loop structure, both the PCV and the BFDV resemble the geminiviruses more than the other circoviruses.

Fig. 5. Comparison of the stem-loop structures of the structurally conserved elements (SCE) found in the ambisense circovirus and representative monocomponent geminivirus genomic DNAs. The viral DNA strand sequences are presented together with the aligned loop sequences. The SLCV loop sequence is highlighted (*). ORF initiation codons are underlined. Iterated sequences (iterons) are underlined according to their polarity ($\longrightarrow$ and $\Longrightarrow$ represent different iterated sequences in the same genome). For comparison, numbering of the circovirus DNA sequences begins with the first base in the (G/T)TAGTATTA loop sequence. This differs from the PCV loop sequence numbering systems reported previously [9, 40]. The inferred sequences of the first few amino acids encoded by the C1 and V1 ORFs are shown below the DNA sequences. Termination codons (#) are indicated

The plant circovirus and CAV genomes are organized such that proteins are encoded on only the encapsidated DNA strands. Analyses of RNA transcripts for the CAV, CFDV and BBTV confirm the production of messenger RNA from only the viral DNA strand [4, 41, 49].

The alignment of the putative circovirus replication associated protein (Rep) sequences (Fig. 3) presents a high degree of similarity among the three plant circoviruses and the two ambisense animal circoviruses (BFDV and PCV). Inferred phylogenetic relationships among the circoviruses and other viruses which replicate via a rolling circle mechanism (representative geminiviruses and bacteriophage ΦX-174) suggest that this group of five viruses (BFDV, PCV, BBTV, CFDV and SCSV) encode Rep proteins originating from a common ancestor [48, 58]. Figure 6 portrays the relationships among the various Rep protein molecules, resulting from a parsimony-based (SEQBOOT, PROTPARS, CONSENSE) analysis, although an essentially identical relationship is obtained using a distance matrix method (SEQBOOT, PROTDIST, FITCH) [21]. These Rep protein sequences differ from those of the geminiviruses, which group into the monocot-infecting and dicot-infecting species as previously reported [48, 58].

An interesting feature of these circovirus Rep protein sequences is their variation from the gemniviruses and the reported consensus for all such proteins [28] in the sequence of motif 2 (see Fig. 3). This histidine pair is strictly conserved in motif 2 across the archebacteria, eubacteria, geminiviruses, and parvoviruses. This pair of histidine residues, with the first immediately following a proline residue, has been implicated in the binding of divalent metal ion cofactors for a number of proteins [11] and is present in that motif in the CAV Rep protein (Fig. 8). However, the remaining circoviruses have diverged and have a substituted glutamime in place of the second histidine in this motif (Fig. 3). It remains to be seen what significance this substitution has on the mechanism of DNA replication in this group of eukaryotic replicons.

Recently, the essential role of the ATP/GTP binding motif (P-loop) in geminivirus DNA replication has been described [18]. Since this P-loop motif is conserved among the circoviruses whose Rep proteins possess it, this virally encoded enzymatic activity is presumably essential for virus replication. The geminivirus Rep protein is known to mediate the cleavage and joining reactions through the three motifs in the amino half of the molecule and is believed to require an ATP-dependent bending of the double stranded replication form of the viral DNA [25, 34]. Thus the ATP binding domain is postulated to act as a topoisomerase or helicase in this bending and local unwinding reaction.

The amino termini of the circovirus capsid proteins possess a high percentage of the basic amino acids arginine and lysine, like the coat proteins of the luteoviruses (Fig. 4a). This suggests that the amino terminal regions of these coat/capsid proteins have DNA binding activity and are in contact with the packaged viral DNA in the native virion. This has been demonstrated in a sobemovirus, the southern bean mosaic virus, for which the structure and relationship of the coat protein to the encapsidated viral nucleic acid have been determined by electron density mapping [24]. This DNA binding activity is also suggested by the

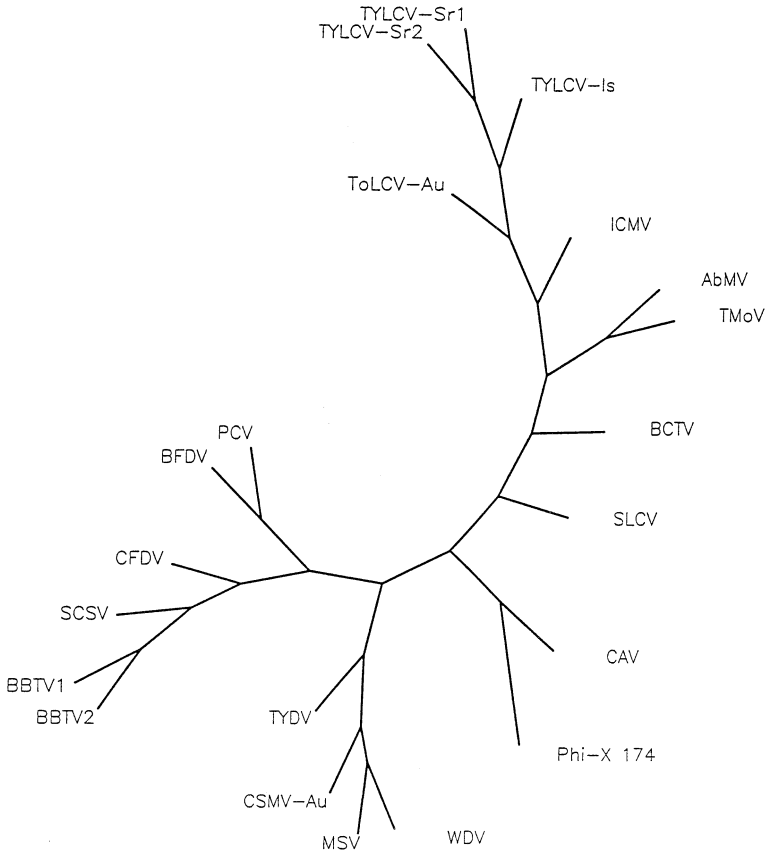


Fig. 6. Unrooted tree depicting the relationships among the Rep proteins for various viruses. Tree is based on parsimony method and was generated using the PHYLIP version 3.5 package (SEQBOOT, PROTPARS, CONSENSE; outgrouped to bacteriophage ΦX-174) from an alignment performed with PILEUP [6]. *AbMV* abutilon mosaic virus; *BBTV1* banana bunchy top virus component 1; *BBTV2* banana bunchy top virus component 2; *BCTV* beet curly top virus; *BFDV* beak and feather disease virus; *CAV* chicken anemia virus; *CFDV* coconut foliar decay virus; *CSMV-Au* Australian isolate of chloris striate mosaic virus; *ICMV* Indian cassava mosaic virus; *MSV* maize streak virus; *PCV* porcine circovirus; *Phi-X 174* bacteriophage ΦX 174; *SCSV* subterranean clover stunt virus; *SLCV* squash leaf curl virus; *TmoV* tomato mottle virus; *ToLCV-Au* Australian isolate of tomato leaf curl virus; *TYDV* tobacco yellow dwarf virus; *TYLCV-SR* Spanish isolates of tomato yellow leaf curl virus; *TYLCV-IS* Israeli isolate of tomato yellow leaf curl virus; *WDV* wheat dwarf virus

high sequence identity between this region of the putative BFDV capsid protein and the highly conserved sequences of the protamines (Fig. 4a [47]).

Differences

The CAV Rep protein is quite different from the other circovirus and geminivirus Rep proteins. It possesses all three sequence motifs found in proteins involved in rolling circle replication. However, unlike the Rep proteins from the other

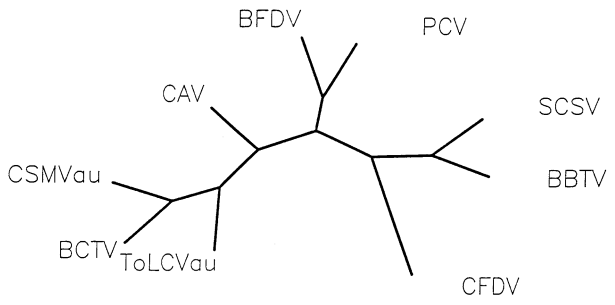


Fig. 7. Unrooted tree depicting the relationships among the coat proteins for the various circoviruses and representative geminiviruses. Tree was constructed as described for Fig. 6 and was outgrouped to the CAV

circoviruses and the geminiviruses it does not have the ATG/GTP binding motif (P-loop; Fig. 8). This protein is quite unique in that it is also the viral capsid protein, possessing the protamine-like basic amino acid sequence at its amino terminus (Fig. 4). This suggests that it has originated quite differently from the other circoviruses, whose Rep proteins share significant sequence identity, and that the CAV falls outside of the inferred circovirus and geminivirus phylogenetic groups (Fig. 6). It also suggests that the CAV replication process relies on a host cell topoisomerase or helicase to provide the local bending and/or unwinding function in the replication mechanism. This use of a host cell topoisomerase/helicase in a rolling circle DNA replication mechanism is known to occur in the case of bacteriophage ΦX-174 and other viruses [28].

With the exception of the basic amino acid residues located at their amino termini, the putative capsid protein sequences are quite divergent. Inferred phylogenetic relationships among these capsid proteins (Fig. 8) group the ambisense-organized circoviruses (BFDV and PCV) together. The two multicomponent plant circovirus capsid proteins (BBTV and SCSV; both aphid-transmitted) also group together. However, the CAV and CFDV capsid proteins may be as different from the other circoviruses as from those of the geminiviruses.

The most obvious differences among these viruses is the coding strategy of the various genomes and the multicomponent nature of two of them. The CAV, CFDV, BBTV, and SCSV appear to have positive strand genomes, encoding proteins from only the encapsidated viral DNA strands (Table 2). Like the geminiviruses, BFDV and PCV possess ambisense genomes, encoding proteins from both the viral DNA strand and a complementary DNA strand that is necessarily present during DNA replication. But they differ from the geminiviruses (Fig. 2g) in that the DNA strand that is packaged in the virion is the strand, which would serve as the transcriptional template for the capsid protein (ORF C1; Fig. 2a and 2b).

Finally, all of the circoviruses except the CAV possess a stem-loop structure like those found in the geminivirus DNAs. No comparable structure can be identified in the CAV noncoding region either based on the near-consensus loop sequence or on the presence of iterated motifs adjacent to any stem-loop structure within this region.

The analysis presented here suggests that the beak and feather disease virus and the porcine circovirus share characteristics of both the plant circoviruses and

Fig. 8. Alignment of CAV VP1 with the putative Rep proteins of the other animal circoviruses. The three conserved motifs for rolling circle replication proteins are shown above the aligned sequences for the BFDV and PCV sequences, and below the aligned sequences for the CAV protein (1, 2, 3; underlined). The P-loop sequence and putative pyrophosphatase domain are highlighted in bold. Letter "u" represents bulky hydrophobic residue [28]

the geminiviruses. The organization of the encapsidated DNA is essentially the complement to that encapsidated by the geminiviruses. The sequence similarity to the plant circovirus Rep proteins suggests a common origin and places the two ambisense circoviruses in an intermediate position. The sequence numbering system used for this analysis was adopted from that used for the geminiviruses [7]. Such a standardized system (see Materials and methods) should be adopted as a convention for the circoviruses.

The evidence presented here also suggests that the plant circoviruses and the ambisense circoviruses belong in different genera within the *Circoviridae*. The recently described faba bean necrotic yellows virus (FBNYV) and milk vetch dwarf virus (MVDV) have also been included in the proposed nanavirus genus [8]. The strikingly different characteristics of the chicken anemia virus genome place it within a more distant category. This division is also supported by the morphological data [60]. Since it is the type species for the family, however, it probably should occupy a genus separate from the other animal circoviruses.

The relationship of the beak and feather disease virus to the plant members presents another interesting facet. The geographic distribution of these viruses reveals that a common ancestor might have existed in the South Pacific. The coconut foliar decay virus is found in Vanuatu [55]. The subterranean clover stunt virus is found in Australia [12]. The banana bunchy top virus is believed to have originated in Fiji (it now has an almost worldwide tropical distribution) [17, 29, 63, 64]. The beak and feather disease virus was described primarily as a problem for aviculture and companion bird owners. Interestingly, the highest reported incidence of disease was in psittacine birds native to the South Pacific islands and Australia [54] and has only recently been demonstrated to infect the New World Psittaciformes [23]. This disease is endemic in free-ranging populations of cockatoos and other psittacines in the South Pacific, and descriptive reports of the disease date to the turn of the centruy [50, 54].

Although serological evidence has suggested the widespread presence of the porcine circovirus in herds in Europe, Canada, and the United States, only recently has the virus been linked to disease states in pigs [14, 15, 26, 43]. Other reports suggest the presence of circoviruses in pigeons, finches, mice, cattle, and perhaps even man [56, 57, 59, 62].

Based on this analysis we propose the following taxonomy for these viruses. First, the family *Circoviridae* with PCV as the type species and including BFDV. These viruses possess non-geminate, non-enveloped virions of 14–20 nm diameter; circular, ssDNA nucleic acids with nonsegmented, ambisense genome organizations and characteristic geminivirus-like stem-loop and repeated sequences in the intergenic region. Second, a new fmily including the BBTV-like viruses (BBTV, SCSV, FBNYV, CFDV) with BBTV as the type species. These viruses possess non-enveloped virions of 17–20 nm diameter; circular, ssDNA nucleic acids with segmented or non-segmented genomes with characteristic geminivirus-like stem-loop in the non-coding regions. Third, the CAV is designated an unassigned virus with circular, ssDNA.

Acknowledgements

This work was supported by the International Avian Research Foundation, Cowan Animal Health Foundation, Association of Avian Veterinarians, American Federation of Aviculture, Terry Clyne, and numerous bird clubs. We are grateful to Dr. M. H. Vodkin, Dr. T. G. Clark, and Dr. J. A. Kimpel for critical discussions and helpful suggestions. We thank D. Pesti, R. P. Campagnoli, Dr. C. M. Otto, and C. J. Leithner for technical assistance.

References

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410
2. Ambrose BJB, Pless RC (1987) DNA sequencing: Chemical methods. Methods Immunol 152: 522–538
3. Arguello-Astorga, GR, Guevara-Gonsalez RG, Herrera-Estrella LR, Rivera-Bustamante RF (1994) Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. Virology 203: 90–100
4. Beetham PR, Hafner GJ, Harding RM, Dale JL (1997) Two mRNAs are transcribed from banana bunchy top virus DNA-1. J Gen Virol 78: 229–236
5. Boevink P, Chu PWG, Keese P (1995) Sequence of subterranean clover stunt virus DNA: affinities with the geminiviruses. Virology 207: 354 –361
6. Bork P, Koonin EV (1994) A p-loop-like motif in a widespread ATP pyrophosphatase domain: implications for the evolution of sequence motifs and enzyme activity. Proteins Struct Funct Genet 20: 347–355
7. Briddon RW, Markham PG (1995) The Geminiviridae. In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD (eds.) Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien New York, pp 158–165 (Arch Virol [Suppl] 10)
8. Brunt A, Crabtree K, Dallwitz M, Gibbs A, Watson L (1996) Viruses of plants: descriptions and lists from the VIDE database. CAB International, Wallingford (URL: http: //biology.anu.edu.au/research-groups/MES/vide/refs.htm)
9. Buhk HJ, Mankertz A (1996) Sequence analysis of porcine circovirus. Unpublished (GenBank Accession no. Y09921)
10. Burns TM, Harding RM, Dale JL (1995) The genome organization of banana bunchy top virus: analysis of six ssDNA components. J Gen Virol 76: 1 471–1 482.
11. Chakrabarti P (1990) Geometry of interaction of metal ions with histidine residues in protein structures. Protein Eng. 4: 57–63
12. Chu PWG, Helms K (1998) Novel virus-like particles containing circular single-stranded DNAs associated with subterranean clover stunt disease. Virology 167: 38–49
13. Chu PWG, Keese P, Qiu BS, Waterhouse PM, Gerlach WL (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 Res 27: 161–171
14. Clark CG (1997) Post-weaning multisystemic wasting syndrome. In: Proceedings of the American Association of Swine Practitioners, Quebec City, Canada, pp 499–501
15. Daft B, Nordhausen RW, Latimer KS, Niagro FD (1996) Interstitial pneumonia and lymphadenopathy associated with circoviral infection in a 6-week old pig. In: Proceedings of the Annual Meeting of the American Association of Veterinary Laboratory Diagnosticians, Little Rock, Arkansas, p 32

16. Dahlhausen B, Radabaugh S (1993) Update on psittacine beak and feather disease and avian polyomavirus testing. Proceedings of the Association of Avian Veterinarians, Nashville, TN, pp 5–7
17. Dale JL (1987) Banana bunchy top: an economically important tropical plant virus disease. Adv Virus Res 33: 301–325
18. Desbiez C, David C, Mettoughi A, Laufs J, Gronenborn B (1995) Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. Proc Natl Acad Sci USA 92: 5 640–5 644
19. Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395
20. Douglas AJ, Phenix K, Mawhinney KA, Todd D, Mackie DP, Curran WL (1995) Identification of a 24 kDa protein expressed by chicken anemia virus. J Gen Virol 76: 1 557–1 562
21. Felsenstein J (1993) PHYLIP version 3.5c, Univ. of Washington, Seattle, WA
22. Genetics Computer Group (1994) Manual for the Wisconsin Package, version 8. Genetics Computer Group, Madison, WI.
23. Greenacre CB, Latimer KS, Niagro FD, Campagnoli RP, D. Pesti D, Ritchie BW (1992) Psittacine beak and feather disease in a scarlet macaw (*Ara macao*). J Assoc Avian Vet 6: 95–98
24. Hermodson M, Abad-Zapatero C, Abdel-Meguid SS, Pundak S, Rossmann MG, Tremaine JH (1982) Amino acid sequence of southern bean mosaic virus coat protein and its relation to the three-dimensional structure of the virus. Virology 119: 133–149
25. Heyraud-Nitschke F, Schumacher S, Laufs J, Schaefer S, Schell J, Gronenborn B (1995) Determination of the origin cleavage and joining domain of geminivirus Rep proteins. Nucleic Acids Res 23: 910–916
26. Hines RK, Lukert PD (1997) Porcine circovirus as a cause of congential tremors in newborn pigs. In: Proceedings of the American Association of Swine Practitioners, Quebec City, Canada, p 344
27. Hodgman TC (1988) A new superfamily of replicative proteins. Nature 333: 22–23
28. Ilyina TV, Koonin EV (1992) Conserved sequence motifs in the initiator proteins for rolling circle replication encoded by diverse replicons from eubacteria, eucaryotes and archebacteria. Nucleic Acids Res 20: 3 279–3 285
29. Karan M, Harding RM, Dale JL (1994) Evidence for two groups of banana bunchy top virus. J Gen Virol 75: 3 541–3 546
30. Katul L, Maiss E, Vetten HJ (1995). Sequence analysis of faba bean necrotic yellows virus DNA component containing a putative replicase gene. J Gen Virol 76: 475–479
31. Katul L, Maiss E, Morozov SY, Vetten HJ (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
32. Koch G, van Roozelaar DJ, Verschueren CAJ, van der Eb AJ, Noteborn MHM (1995) Immunogenic and protective properties of chicken anemia virus proteins expressed by baculovirus. Vaccine 8: 763–770
33. Latimer KS, Niagro FD, Rakich PM, Campagnoli RP, Ritchie BW, Steffens WL, Pesti D, Lukert PD (1992) Comparison of DNA dot-blot hybridization, immunoperoxidase staining and routine histopathology in the diagnosis of psittacine beak and feather disease in paraffin-embedded cutaneous tissues. J Assoc Avian Vet 6: 165–168
34. Laufs J, Traut W, Heyraud F, Matzeit V, Rogers SG, 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. Proc Natl Acad Sci USA 92: 3 879–3 883
35. Lukert PD, de Boer GF, Dale JL, Keese P, McNulty MS, Randles JW, Tisher I (1995) The Circoviridae. In: Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW,

Martelli GP, Mayo MA, Summers MD (eds) Virus Taxonomy. Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Springer, Wien New York, pp 166–168 (Arch Virol [Suppl] 10)
36. Mankertz J, Buhk HJ (1990) Transcriptional analysis of porcine circovirus (PCV). VIIIth International Congress of Virology, Abstracts, p 380
37. Mankertz A, Persson F, Mankertz J, Blaess G, Buhk HJ (1997) Mapping and characterization of the origin of DNA replication of porcine circovirus. J Virol 71: 2 562–2 566
38. Mayo MA, Robinson DJ, Jolly CA, Hyman L (1989) Nucleotide sequence of potato leafroll luteovirus RNA. J Gen Virol 70; 1 037–1 051
39. Meehan BM, Todd D, Creelan JL, Earle JAP, Hoey EM, McNulty MS (1992) Characterization of viral DNAs from cells infected with chicken anemia agent: sequence analysis of the cloned replicative form and transfection capabilities of cloned genome fragments. Arch Virol 124: 301–319
40. Meehan BM, Creelan JL, McNulty MS, Todd D (1997) Sequence of porcine circovirus DNA: affinities with plant circoviruses. J Gen Virol 78: 221–227
41. Merits A, Zelenina DA, Mizenina OA, Chernov BK, Morozov SY (1995) Poly(A) addition site mapping and polyadenylation signal analysis in a plant circovirus replication-related gene. Virology 211: 345–349
42. Myers EW, Miller W (1988) Optimal alignments in linear space. CABIOS 4: 11–17
43. Nayar GP, Hamel A, Lin L (1997) Detection and characterization of porcine circovirus associated with post weaning multisystemic wasting syndrome in pigs. Can Vet J 38: 385–386
44. Niagro FD (1993) Unpublished data
45. Noteborn MHM, de Boer GF, van Roozelaar DJ, Karreman C, Kranenburg O, Vos JG, Jeurissen HM, Hoeben RC, Zantema A, Koch G, van Ormondt H, van der Eb AJ (1991) Characterization of cloned chicken anemia virus DNA that contains all elements for the infectious replication cycle. J Virol 65: 3 131–3 139
46. Noteborn MHM, Todd D, Verschueren CAJ, de Gauw HWFM, Curran WL, Veldkamp S, Douglas AJ, McNulty MS, van der Eb AJ, Koch G (1994) A single chicken anemia virus protein causes apoptosis. J Virol 68: 346–351
47. Oliva R, Goren R, Dixon GH (1989) Quail (*Cotournix japonica*) protamine, full-length cDNA sequence, and the function and evolution of vertebrate protamines. J Biol Chem 264: 17 627–17 630
48. Padidam M, Beachy RN, Fauquet C (1995) Classification and identification of geminiviruses using sequence comparisons. J Gen Virol 76: 249–263
49. Phenix KV, Meehan BL, Todd D, McNulty MS (1994) Transcriptional analysis and genome expression of chicken anemia virus. J Gen Virol 75: 905–909
50. Raidal SR, McElenea CL, Cross GM (1993) Seroprevalence of psittacine beak and feather disease in wild psittacine birds in new south wales. Aust Vet J 70: 137–139
51. Ramis A, Latimer KS, Niagro FD, Campagnoli RP, Ritchie BW, Pesti D (1994) Diagnosis of psittacine beak and feather disease (PBFD) viral infection, avian polyomavirus infection, adenovirus infection and herpesvirus infection in psittacine tissues using DNA in situ hybridization. Avian Pathol 23: 643–657
52. Randles JW, Hanold D (1989) Coconut foliar decay virus particles are 20-nm icosahedra. Intervirology 30: 177–180
53. Ritchie BW, Niagro FD, Lukert PD, Steffens WL, Latimer KS (1989) Characterization of a new virus from cockatoos with psittacine beak and feather disease. Virology 171: 83–88
54. Ritchie BW (1995) Avian viruses: function and control. Wingers Publishing, Lake Worth

55. Rohde W, Randles JW, Langridge P, Hanold D (1990). Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. Virology 176: 648–651
56. Rohwer RG (1991) The scrapie agent: "a virus by any other name". Curr Top Microbiol Immunol 172: 195–231
57. Shivaprasad HL, Chin RP, Jeffrey JS, Latimer KS, Nordhausen RW, Niagro FD, Campagnoli RP (1994) Particles resembling circovirus in the bursa of fabricius of pigeons. Avian Dis 38: 635–641
58. Timmermans MCP, Das OP, Messing J (1994) Geminiviruses and their uses as extrachromosomal replicons. Ann Rev Plant Physiol Plant Mol Biol 45: 79–112
59. Tischer I, Bode L, Apodaca J, Timm H, Rasch R, Pociulli S, Gerike E (1995) Presence of antibodies reacting with porcine circovirus in sera of humans, mice, and cattle. Arch Virol 140: 1 427–1 439
60. Todd DL, Niagro FD, Ritchie BW, Curran W, Allan GM, Lukert PD, Latimer KS, Steffens WL, McNulty MS (1991) Comparison of three animal viruses with circular single-stranded DNA genomes. Arch Virol 117: 129–135
61. Wilson ME, Mainprize TH, Friesen PS, Miller LK (1987) Location, transcription, and sequence of a baculovirus gene encoding a small arginine-rich polyprotein. J Virol 61: 661–666
62. Woods LW, Latimer KS, Niagro FD, Riddle C, Crowley AM, Anderson ML, Daft BM, Moore JD, Campagnoli RP, Nordhausen RW (1994) A retrospective study of circovirus infection in pigeons: nine cases (1986–1993). J Vet Diagn Invest 6: 156–164
63. Xie WS, Hu JS (1995) Molecular cloning, sequence analysis, and detection of a banana bunchy top virus in Hawaii. Phytopathology 85: 339–347
64. Yeh HH, Su HJ, Chao YC (1994) Genome characterization and identification of viral-associated dsDNA component of banana bunchy top virus. Virology 198: 645–652

Authors' address: Dr. F. Niagro, Clinical Investigation Division, Dwight D. Eisenhower Army Medical Center, Fort Gordon, GA 30 905–5 650, U.S.A.

Received March 3, 1998