Nucleotide sequence analysis of a novel circovirus of canaries and its relationship to other members of the genus Circovirus of the family Circoviridae

Kerry V. Phenix,1 Jonathan H. Weston,2 Ingrid Ypelaar,1 Antonio Lavazza,3 Joan A. Smyth,4 Daniel Todd,4 Graham E. Wilcox1 and Shane R. Raidal1

1 Division of Veterinary and Biomedical Sciences, Murdoch University, Murdoch, Western Australia 6150, Australia
2 Department of Veterinary Science, The Queen's University of Belfast, Stormont, Belfast BT4 3SD, Northern Ireland
3 Electron Microscopy Laboratory, Istituto Zooprofilattico Sperimentale della Lombardia e dell’Emilia Romagna, Via Bianchi 9, 25124 Brescia, Italy
4 Department of Agriculture and Rural Development for Northern Ireland, Stormont, Belfast BT4 3SD, Northern Ireland

The circular, single-stranded DNA genome of a novel circovirus of canaries, tentatively named canary circovirus (CaCV), was cloned and sequenced. Sequence analysis indicated that the genome was 1952 nucleotides (nt) in size and had the potential to encode three viral proteins, including the putative capsid and replication-associated (Rep) proteins. The CaCV genome shared greatest sequence similarity (58.3% nt identity) with the newly characterized columbid circovirus (CoCV) and was more distantly related to the two porcine circovirus strains, PCV1 and PCV2, beak and feather disease virus (BFDV) and a recently isolated goose circovirus (GCV) isolate (46.8–50.9% nt identity). In common with other members of the Circovirus genus, several nt structures and amino acid motifs thought to be implicated in virus replication were identified on the putative viral strand. Phylogenetic analysis of both the capsid and Rep protein-coding regions provided further evidence that CaCV is more closely related to CoCV and BFDV and more distantly related to GCV, PCV1 and PCV2.

A growing number of circovirus and circovirus-like infections of avian species have recently been described (Todd, 2000; Woods & Latimer, 2000). Currently, the Circoviridae comprises four species, chicken anaemia virus (CAV; Noteborn et al., 1991), porcine circovirus PCV1 (Meehan et al., 1997) and PCV2 (Meehan et al., 1998) and psittacine beak and feather disease virus (BFDV; Bassami et al., 1998; Niagro et al., 1998), all of which possess small, circular, single-stranded DNA genomes. On the basis of its unique genome organization, CAV has recently been re-classified as the only member of the genus Gyrovirus (Pringle, 1999), and PCV1, PCV2 and BFDV were classified as members of the genus Circovirus. Sequence analysis of the genomes of two additional, tentative members of the circovirus family, columbid circovirus (CoCV) and goose circovirus (GCV), indicates that these viruses should also be assigned to the genus Circovirus (Mankertz et al., 2000; Todd et al., 2001a).

Circovirus infections are commonly associated with immunodeficiency-related diseases that are potentially fatal. A condition known as ‘black spot’ of neonatal canaries, characterized by abdominal enlargement, gall bladder congestion and failure to thrive, has been described for many years in Europe and was reported to have been caused by a circovirus (Goldsmith, 1995). More recently, a circovirus-like infection of canaries was identified in adult birds that had died following a short illness characterized by dullness, anorexia, lethargy and feather disorder (Todd et al., 2001b). Large numbers of circovirus-like, spherical virus particles were detected by electron microscopy in the organ homogenates of diseased birds. A degenerate primer-based PCR technique using degenerate primers specific to conserved amino acid (aa) sequences in the circovirus Rep (replication-associated) protein successfully amplified a circovirus (CaCV)-specific DNA fragment from infected birds (Todd et al., 2001b). Nucleotide (nt) and predicted aa sequence analysis of the 510 bp genomic fragment indicated that the CaCV sequence was more closely related to CoCV than to BFDV, and more distantly related to PCV1 and PCV2. Here we report the complete genome sequence of CaCV and discuss its relationship with other members of the Circoviridae.

The full-length circular CaCV genome was amplified by inverse PCR (IP–PCR) from the original CaCV clone described previously (Todd et al., 2001b) using the primers CaCV1 5′

Author for correspondence: Shane Raidal.
Fax +61 93104144. e-mail raidal@numbat.murdoch.edu.au
ATGATTGGAGACAGAAACCCCGTGAC 3' and CaCV2 5' GTCCAATCATCAGGGCCAAATCCC 3'. PCR was carried out at 94 °C for 30 s, 60 °C for 30 s, 72 °C for 2 min for 35 cycles using the Expand Long Template PCR system (Roche Diagnostics) as described by the manufacturer. The resulting IP–PCR product (approximately 2 kb) was cloned into the pCR2.1-TOPO vector (Invitrogen) and two clones derived from the same IP–PCR product were selected for sequencing. Clones were sequenced in triplicate in both directions. The nt sequence of CaCV was edited and assembled using SeqEd version 1.0.3 (Applied Biosystems) and analysed using a range of programs provided by the Australian National Genomic Information Services (ANGIS). For multiple sequence alignment and phylogenetic analysis of sequences the programs Global Pair Alignment, Gap, PileUp and ClusTree were used. Sequence data were edited and aligned by the PileUp program and analysed phylogenetically using the ClusTree program with 1000 bootstrap cycles. ClusTree computes a phylogenetic tree according to the neighbour-joining method of Saitou & Nei (1987). Database searches were performed in non-redundant nucleic and protein databases with the BLAST program (Altschul et al., 1990). The GenBank accession number for the CaCV sequence is AF346618.

Sequence analysis revealed that the genome of CaCV was circular and had a size of 1952 nt. The nt numbering and open reading frame (ORF) nomenclature used was similar to that adopted previously for animal circoviruses (Todd et al., 2000); nt position ‘1’ is given as the ‘A’ residue at position 8 of a nonamer sequence (CAGTATTAC) present at the apex of a potential stem–loop structure. This stem–loop/nonamer element is conserved in circoviruses, with the exception of CAV in which only the nonamer is present, and is thought to represent the origin of rolling circle replication. In CaCV, the nonamer sequence differs slightly from that of other circoviruses in that the first residue of the nonamer is a ‘C’, rather than the ‘T’ residue found in other members of the genus. Such variations in the nonamer sequence (Meehan et al., 1997, 1998; Bassami et al., 1998) and also in the number of base-pair residues comprising the stem–loop structure (Fig. 1A, B) have previously been reported (Todd et al., 2001a). In addition, three base differences at nt positions 1825 (C or T), 1931 (C or T) and 1945 (A or G) were found between the two clones sequenced.

In common with other circoviruses, the stem–loop structure is found within a putative intergenic region, 77 nt in size, located between the starting codons of the two major ORFs (nt position 1907–31). Close examination of this region identified a series of direct and inverted repeat sequences. The sequence 5’ CCAAAGTG GCCG 3’ forms an inverted repeat (nt position 1932–1942 and 13–3), and constitutes part of the stem–loop structure. In addition, two direct tandem repeats with the sequence 5’ GGAGGCCAC 3’ (nt position 12–19 and 20–27)
were identified adjacent to the stem-loop (Fig. 1A). This set of tandemly repeated octamers was also located within the intergenic region of CoCV (Todd et al., 2001a). Furthermore, Mankertz et al. (2000) described the occurrence of the hexamer sequence 5′ GGAGCC 3′ occurring four times in or adjacent to the stem-loop structure. It has been suggested that such repeat
sequences may play a role in virus replication, perhaps by acting as potential binding sites for the circovirus Rep protein (Mankertz et al., 1997, 2000; Niagro et al., 1998; Todd et al., 2001).

Genome analysis indicated that CaCV displays an ambisense genome organization, with one large ORF being encoded on the viral strand and one major and one minor ORF on the complementary strand (Fig. 1C). The largest ORF, V1 (290 aa), encodes the putative Rep protein, with a predicted molecular mass of 33.4 kDa. Multiple alignment of the deduced aa sequence of the Rep protein of CaCV with those encoded by other circoviruses including PCV1, PCV2, BFDV, CoCV and GCV, identified conserved aa motifs typical for rolling circle replication and also the dNTP-binding domain. Three additional conserved motifs, WWDGY, DDFYGWLP and DRYP, previously reported to be present in CoCV and GCV (Todd et al., 2001a), were also identified within the Rep protein of CaCV (Fig. 2). A putative poly(A) signal (AAATAAA) was also identified on the viral strand, 39 nt (nt position 943–948) downstream from the ORF V1 stop codon.

The complementary strand of CaCV had one major ORF, ORF C1 (250 aa), encoding the putative capsid protein (molecular mass of 30 kDa) and one minor ORF, ORF C2 (134 aa), encoding a protein with a predicted molecular mass of 14.9 kDa. Multiple sequence alignment of the deduced aa sequence of ORF C1 revealed a highly basic N-terminal region, similar to that found in the putative capsid proteins of other circoviruses. In addition, a number of minor conserved motifs were identified in the C1 ORFs of the avian circoviruses, but not in GCV ORF C1 (data not shown). ORF C2 appeared to use an alternative TCT start codon, located at nt position 679. This usage of alternative start codons in circovirus protein translation has previously been described for BFDV (Bassami et al., 2001; Niagro et al., 1998; Mankertz et al., 2000) and GCV (Todd et al., 2001a). The function, if any, of the putative ORF C2 protein is unknown. It is similar in size to the ORF C2 of CoCV (13.7 kDa) and the deduced aa sequence shows similarity to both the CoCV ORF C2 (44%) and the BFDV ORF 3 (16%) proteins by GAP analysis. A candidate poly(A) signal was identified for the C1 ORF at nt position 1158–1153.

In terms of the relationship of CaCV with other circoviruses, nt and aa sequence comparisons indicate that CaCV is more closely related to CoCV and BFDV than to the recently characterized avian circovirus GCV and the two porcine circoviruses, PCV1 and PCV2. Pairwise comparisons indicated that CaCV showed greatest homology to CoCV (58.3% nt identity) and BFDV (55.7% nt identity) and less homology to GCV (49.5% nt identity), PCV1 (50.9% nt identity) and PCV2 (46.8% nt identity).

As expected, highest levels of aa identity were shared between the Rep-coding regions of circoviruses. Pairwise comparisons showed that the level of aa identity of the V1 ORF shared by CaCV with CoCV (63.4%) and BFDV (62.9%) was substantially higher than that shared by CaCV with GCV (45.8%), PCV1 (42.5%) and PCV2 (40.4%). Overall, less homology was observed between the capsid protein-coding regions, with CaCV ORF C1 showing greatest aa identity with the C1 ORFs of BFDV (44%) and CoCV (39%). This level of aa identity was considerably reduced when the C1 ORFs of CaCV were compared with PCV1 (28.5%), PCV2 (29.8%) and GCV (25.7%). Phylogenetic trees derived from the predicted aa sequences of the Rep and capsid proteins of the six members of the Circovirus genus demonstrate these relationships (Fig. 3).

On the basis of nt and aa sequence homology, we recommend that CaCV be included as the latest member of the genus Circovirus of the family Circoviridae. It is hoped that the additional information provided by this work will aid the development of diagnostic tests, based on the detection of viral DNA, with which to assess the prevalence of CaCV and associated disease.

References


Received 15 May 2001; Accepted 7 August 2001