INTRODUCTION

Members of the family Circoviridae are non-enveloped icosahedral particles with a diameter of 17–26.5 nm and a circular single-stranded DNA genome, approx. 2 kb in size (Todd et al., 2005). The family comprises the genera Gyrovirus and Circovirus; in the latter the genome is organized in an ambisense direction. The replication-associated (Rep) protein is encoded from the virus sense-strand (ORF-V1) and the capsid protein is encoded from the complementary sense-strand (ORF-C1) of the genomes of all viruses of the genus Circovirus; additional small open reading frames (ORFs) are present in some of the circoviruses (Niagro et al., 1998; Todd et al., 2001b).

Until now, seven circoviruses have been described that fit the criteria to be placed into the genus Circovirus. The avian circoviruses Beak and feather disease virus (BFDV, Niagro et al., 1998; Bassami et al., 1998), pigeon circovirus (PiCV, Mankertz et al., 2000; Todd et al., 2001b), canary circovirus (CaCV, Phenix et al., 2001), goose circovirus (GoCV, Todd et al., 2001b) and duck circovirus (DuCV) isolates from Muscovy duck (Hattermann et al., 2003) and Mulard duck (Soike et al., 2004) are involved in diseases of birds, with the main clinical symptoms including immunosuppression and feather disorders in young birds (Todd, 2004; Raue et al., 2005). Among the two porcine circoviruses (PCV), PCV-2 causes the post-weaning multisystemic wasting syndrome and is associated with porcine dermatitis and nephropathy syndrome (Chae, 2005), whereas PCV-1 could not be associated with overt disease (Allan & Ellis, 2000).

The novel circoviruses were mainly detected by degenerate primer PCR and the remaining genome was amplified by inverse PCR (Todd et al., 2001a, b; Hattermann et al., 2003). This technique, however, requires the existence of highly conserved target sequences for primer binding in the genomes of the unknown viruses. Recently, the technique of multiply primed rolling-circle amplification (RCA) has been successfully used for amplification of the circular genomes of papillomaviruses (Rector et al., 2004a, 2005) and anelloviruses (Niel et al., 2005). This technique uses the polymerase of bacteriophage φ29 for selective amplification of circular DNA (Dean et al., 2001). A random hexamer primer anneals to multiple sites of the template DNA, which are isothermally extended by the φ29 DNA polymerase. By strand displacement synthesis, repeated copies of the complete genome are synthesized, leading to a high molecular mass double stranded DNA, from which single genome units can be produced using a single cutting restriction enzyme. As the technique does not need any specific primer, the genus Circovirus comprises small non-enveloped viruses with a circular single-stranded DNA genome. By using PCR with degenerate primers, a novel circovirus (starling circovirus, StCV) was detected in spleen samples of wild starlings (Sturnus vulgaris and Sturnus unicolor) found dead during an epidemic outbreak of septicæmic salmonellosis in northeastern Spain. Using a specific PCR, StCV was also detected in apparently healthy birds from the same population. The genome was amplified using multiply primed rolling-circle amplification and cloned. Open reading frames (ORFs) with similarities to the replication-associated protein and the capsid protein of circoviruses as well as an additional ORF encoding a protein of 106 aa were evident from the sequence. Phylogenetic analysis of circovirus genomes revealed the highest degree of similarity (67–1%) between StCV and canary circovirus. A similar analysis of the evolutionarily conserved cytochrome b gene of the circovirus host species revealed a strict co-evolution of circoviruses with their hosts; however, the circoviruses showed about a threefold higher genetic divergence than their hosts.

The GenBank/EMBL/DDBJ accession number of the sequence reported in this paper is DQ172906.

Genome of a novel circovirus of starlings, amplified by multiply primed rolling-circle amplification

Reimar Johne,¹ Daniel Fernández-de-Luco,² Ursula Höfle³ and Hermann Müller¹

¹Institute for Virology, Faculty of Veterinary Medicine, University of Leipzig, An den Tierkliniken 29, D-04103 Leipzig, Germany
²Department of Animal Pathology, Veterinary Faculty, University of Zaragoza, Spain
³Centro de Investigación Agropecuaria, El Deheson del Encinar, Oropesa, Spain

The novel circovirus was amplified using multiply primed rolling-circle amplification and cloned. Open reading frames (ORFs) with similarities to the replication-associated protein and the capsid protein of circoviruses as well as an additional ORF encoding a protein of 106 amino acids were evident from the sequence. Phylogenetic analysis of circovirus genomes revealed the highest degree of similarity (67–1%) between StCV and canary circovirus. A similar analysis of the evolutionarily conserved cytochrome b gene of the circovirus host species revealed a strict co-evolution of circoviruses with their hosts; however, the circoviruses showed about a threefold higher genetic divergence than their hosts.
RCA is convenient for amplification of any circular DNA including a majority of viral genomes (Rector et al., 2004b).

In this paper, the detection of a novel circovirus in wild European starlings (Sturnus vulgaris) and spotless starlings (Sturnus unicolor), and the application of RCA for the amplification of its genome is described. The genome sequence was analysed and phylogenetic relationships were determined. Aspects of virus-host evolution of circoviruses as well as the clinical significance of the StCV infection are discussed.

METHODS

Bird samples. Bird samples were collected during a mortality incident with approximately 80,000 deaths in the wild population of starlings in northeastern Spain in January–February 2005. After necropsy, organ samples were subjected to histopathological and microbiological examination. In all of the diseased birds tested, Salmonella species had been detected in many organs (Arnal et al., 2005). Organ samples from wild-caught apparently healthy birds from the same region were negative for Salmonella species. Spleen samples from diseased as well as apparently healthy European starlings (Sturnus vulgaris), spotless starlings (Sturnus unicolor) and Eurasian jackdaws (Corvus monedula) were analysed by PCR. The bursa of Fabricius of a pigeon previously tested positive for PiCV (Raue et al., 2005) was used in control experiments.

Degenerate primer PCR for circovirus detection. DNA was isolated from the samples by using the DNeasy tissue kit (Qiagen). PCR for detection of circovirus DNA was carried out as described by Todd et al. (2001a), with minor modifications. Briefly, the primers 5′-TTACCCTTAAYAYCCT-3′ and 5′-CCTATATCCACCCA-3′ were used in PCR with Taq DNA polymerase (Peqlab) with the following cycling profile: 95°C for 5 min, 40 cycles with 94°C for 30 s, 52°C for 30 s and 72°C for 1 min, and a final incubation for 5 min at 72°C. PCR products were analysed by electrophoresis on ethidium bromide-stained 2% agarose gels and the vector PCR-TOPO (Invitrogen). The insert was sequenced using the primer M13 in an ABI Prism device (Applied Biosystems). The sequence was checked for similarity with other DNA sequences using the BLAST search facility (Altschul et al., 1997).

PCR for detection of StCV genome. The PCR specific for the starling circovirus (StCV) was performed using primers delineated from a previously determined partial nucleotide sequence of the ORF-V1 (5′-AAAGAGCTCAGGGAAGAAGGC-3′ and 5′-CCCCG-CCCCATATCTACCTAAG-3′) with Taq DNA polymerase (Peqlab), buffer Y (Peqlab) and DNA isolated as above as template. The cycling profile consisted of an incubation at 95°C for 5 min, 40 cycles with 94°C for 30 s, 60°C for 30 s and 72°C for 30 s, followed by a final incubation at 72°C for 5 min. The PCR product with a length of 208 bp was analysed as above.

Multiply primed RCA and cloning of the StCV genome. RCA was performed as described by Rector et al. (2004b) using the TempliPhi 100 amplification kit (Amersham Biosciences). Briefly, 1 μl DNA was isolated as above and mixed with 5 μl TempliPhi sample buffer supplemented with 450 μM extra dNTPs, the sample was incubated at 95°C for 3 min and subsequently cooled on ice. After adding 5 μl TempliPhi reaction buffer and 0.2 μl TempliPhi enzyme mix, the mixture was incubated at 30°C for 16 h, and thereafter inactivated at 65°C for 10 min. Restriction enzyme analysis was performed by using EcoRI with 2 μl of the mixture. For cloning, a total of 30 μl RCA reaction was digested with EcoRI and a resulting 2-kbp fragment was ligated with the EcoRI-restricted vector pBluescript II SK(+) (Stratagene) and transformed into XL-1 Blue MRF′ Escherichia coli cells (Stratagene).

DNA sequencing and analysis of genome sequences. The insert of the plasmid containing the 2-kbp fragment was sequenced using the primers M13 forward and M13 reverse (Invitrogen) in an ABI Prism device (Applied Biosystems). The remaining part was sequenced using the primer-walking method covering both strands of the whole insert. The genome sequence of StCV was reassembled from the sequence fragments using the EditSeq module of the DNASTAR software package (Lasergene) and the complete sequence has been deposited in the GenBank database (accession no. DQ172906).

Sequence alignments and construction of phylogenetic trees were carried out with the CLUSTAL W method (Thompson et al., 1994) using the MegaAlign module of the above-mentioned software package. Bootstrap analysis of the phylogenetic trees was performed using the T-Rex 4.0a1 software (Vladimir Makarenkov, University of Quebec, Montreal) with the neighbour-joining method and 100 replicate calculations. The following sequence data were used (GenBank accession nos in parentheses) for circovirus genomes: CaCV (AJ301633), PiCV (AF252610), BFDV (AF080560), DuCV isolate TC1/2002 from Muscovy duck (in this study designated DuCV-Cm due to the host species Cairina moschata; AJ394721), DuCV isolate from Mulard duck (in this study designated DuCV-Ap due to the host species Anas platyrhynchos; AJ228555), GoCV (AJ304456), PCV-1 (U49186), PCV-2 (AF055392); for cytochrome b: starling (Sturnus vulgaris, AY352540), canary (Serinus canaria, L76266), pigeon (Columba livia, AF182694), cockatoo (Cacatua galerita, AF313735), Muscovy duck (Cairina moschata, L08385), Mulard duck (Anas platyrhynchos, AF059081), goose (Anser anser, AY427802), swine (Sus scrofa, X56295).

RESULTS

Detection of a novel circovirus in the spleen of starlings

Starlings and jackdaws found dead during an epidemic outbreak in northeastern Spain showed multiple haemorrhages and necroses in the intestines, liver and spleen. Salmonella species were isolated from all of the diseased birds tested (Arnal et al., 2005). Spleen samples from diseased European starlings (Sturnus vulgaris), spotless starlings (Sturnus unicolor) and Eurasian jackdaws (Corvus monedula) as well as from apparently healthy birds were investigated by PCR using degenerate primers with binding sites in the ORF-V1 of circoviruses. After electrophoresis of PCR products, faint bands with a length of approximately 0.5 kbp were visible in the cases of all starling samples (Fig. 1a). The PCR products of two samples (sample Su1 and Sv1 in Fig. 1) were cloned and sequenced. Alignment of the nucleotide sequences revealed a similarity of 97.6% between both sequences and to the genome sequences of circoviruses (not shown). The novel circovirus was tentatively designated starling circovirus (StCV). Based on the sequence, a specific PCR amplifying a 208 bp fragment of the StCV genome was established. Using this PCR, all starling samples showed bands with the expected length, indicating infection with StCV (Fig. 1b). Notably, a sample of a PiCV-infected pigeon, which reacted positive in the degenerate primer PCR (P in Fig. 1a) was negative in the
StCV-specific PCR (P in Fig. 1b), indicating specificity of the latter PCR protocol.

**Amplification and cloning of the StCV genome**

Attempts to amplify the remaining part of the StCV genome by inverse PCR using several specific primer pairs with binding sites in the determined sequence were not successful (not shown). Therefore, the RCA technique was applied to sample Sv1. After digestion of the RCA product with EcoRI, a band with a length of approximately 2 kbp was visible (Fig. 1c), subsequently cloned and sequenced by the primer-walking method. Sequence analyses revealed that the cloned fragment enclosed the sequence of the PCR product mentioned above, indicating that the StCV genome had been cloned. A PCR amplifying the region including the EcoRI site did not reveal any additional sequences (not shown), indicating that the EcoRI site was unique and that the whole genome was cloned.

To verify the sequence of the cloned StCV genome and to assess the accuracy of the applied technique, a second RCA was performed with sample Sv1 in an independent reaction and the product was cloned and sequenced as above. The sequences of both clones were compared and two point mutations were detected: G at position 1378 was mutated to T without an amino acid exchange and C at position 1682 was mutated to A leading to an exchange of a proline residue with a serine residue in the putative capsid protein, and C at position 1682 was mutated to T without an amino acid exchange.

**Analysis of the StCV genome sequence**

The genome of StCV consists of 2063 nt. Analysis of the assembled circularized sequence (Fig. 2a) revealed two ORFs in opposite direction encoding proteins with similarities to the circovirus Rep and capsid proteins. An additional ORF (ORF-V2) with a coding capacity of 106 aa (predicted molecular mass 11 kDa for the protein) was found starting in the 3’ region of the ORF-V1, oriented in the same direction and partially overlapping with it, but using another reading frame. A search using the BLAST facility revealed no significant similarities with known proteins for the amino acid sequence translated from ORF-V2 (Fig. 2b). As in other circovirus genomes, a non-coding region is situated between ORF-V1 and ORF-C1, which contains inverted repeat sequences capable of generating a stem–loop, 17 bp in size in the case of StCV (Fig. 2c). This sequence also contained the nonamer sequence (consensus sequence 5′-TAGTATTAC-3′), which is highly conserved in all circovirus genomes; in the case of StCV, however, the sequence is slightly modified to 5′-CAGTATTAC-3′ and therefore identical with that of CaCV. According to the other circovirus genomes, position ‘1’ of nucleotide numbering was set at the eighth ‘A’ residue of the nonamer sequence. Two tandem repeats of the sequence 5′-GGAGCCA-3′, which are putative binding sites for the Rep protein (Phenix et al., 2001), are located at positions 12–18 and 20–26. A putative hairpin structure was found in the opposite region of the genome (Fig. 2d), which was difficult to sequence probably due to the formation of a stable secondary structure (not shown).

The putative Rep protein of StCV consists of 289 aa residues; the lengths of the other circovirus Rep proteins are in the range of 289 (BFDV) and 317 (PiCV) aa residues. The StCV Rep protein has a predicted molecular mass of 33·2 kDa. Comparison of the amino acid sequence with the other circovirus Rep sequences revealed similarities between 41·7 and 76·2 %, with the highest degree of similarity to the Rep of CaCV (Table 1). The highly conserved sequences known to be involved in rolling-circle replication and dNTP-binding of circoviruses (Phenix et al., 2001) are all present in the StCV Rep. This is in contrast to the CaCV Rep, in which the first motif FTLNNPNP is mutated to FTLNNNY.
(Phenix et al., 2001). The ORF-C1 encoding the capsid protein of StCV has a coding capacity of 276 aa [other circovirus capsid proteins range between 218 (PCV-2) and 273 (PiCV) aa] for a protein with a predicted molecular mass of 32-5 kDa. Similarities to other circovirus capsid proteins were in the range of 21-6 and 52-4 %; this protein showed the highest degree of similarity to the PiCV capsid protein. As in all circovirus capsid proteins, the N-terminal region is highly basic probably due to an involvement in packaging of the viral genome into the viral capsid (Johne et al., 2004), with an accumulation of 23 arginine residues in the first 46 aa positions of the StCV capsid protein.

An alignment of the whole genome sequence of StCV with those of the other circoviruses revealed similarities between 14-5 and 67-1 %. The highest degree of similarity was observed between StCV and CaCV genome sequences, which is also reflected by a phylogenetic tree established for the genome sequences of the circoviruses (Fig. 3, left). These viruses form a branch together with PiCV and BFDV, which is separated from another branch formed by GoCV and the DuCV isolates. The porcine circoviruses PCV-1 and PCV-2 group in a distinct branch, indicating only distant relation of these viruses to the avian circoviruses. To assess the relationship between the host species of the circoviruses, a 699 bp fragment of their cytochrome b gene was sequenced using the same method as applied to the circovirus genomes. The phylogenetic tree obtained for the hosts (Fig. 3, right) shows a branching pattern very similar to that of the circoviruses (Fig. 3, left), indicating a strict co-evolution of the viruses with their hosts. However, minor differences are obvious as in the alignment of the host sequences pigeon and cockatoo group together with goose and ducks, whereas their viruses (PiCV and BFDV) group together with StCV and CaCV. A bootstrap analysis (numbers in the trees in Fig. 3) shows values lower than 80 % for the knots separating the respective branches of the avian species, which indicates some uncertainty for their grouping based on the available sequence data. Interestingly, the degree of divergence specified as the number of nucleotide substitution events (scale in Fig. 3) is about threefold higher in the case of the viruses (52-8) as compared with their hosts (16-9), indicating a higher mutation rate of the circoviruses.

### Table 1. Comparison of sequence similarities between StCV and other circoviruses

Abbreviations of the circovirus designations are as in Fig. 3.

<table>
<thead>
<tr>
<th></th>
<th>CaCV</th>
<th>PiCV</th>
<th>BFDV</th>
<th>DuCV-Cm</th>
<th>DuCV-Ap</th>
<th>GoCV</th>
<th>PCV-1</th>
<th>PCV-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rep*</td>
<td>76-2</td>
<td>73-1</td>
<td>61-6</td>
<td>46-6</td>
<td>47-2</td>
<td>47-2</td>
<td>44-5</td>
<td>41-7</td>
</tr>
<tr>
<td>Capsid†</td>
<td>51-2</td>
<td>52-4</td>
<td>42-9</td>
<td>24-1</td>
<td>22-6</td>
<td>21-6</td>
<td>24-0</td>
<td>24-3</td>
</tr>
<tr>
<td>Genome‡</td>
<td>67-1</td>
<td>62-4</td>
<td>25-7</td>
<td>19-7</td>
<td>23-5</td>
<td>24-7</td>
<td>21-9</td>
<td>14-5</td>
</tr>
</tbody>
</table>

* Amino acid sequence identity for the ORF-V1 encoding the Rep protein.
† Amino acid sequence identity for the ORF-C1 encoding the capsid protein.
‡ Nucleotide sequence identity for the whole genome sequence.
DISCUSSION

The members of the genus Circovirus are widely distributed, especially in avian species. Until now, five avian circoviruses and two swine circoviruses have been identified with different clinical significance. The avian circoviruses have been associated with immunosuppressive diseases and feather disorders; however, clinically healthy birds also have been found to be infected with circoviruses making interpretation of their aetiological role in diseases difficult (Todd, 2004; Raue et al., 2005). For pigeon circovirus infection, it has been speculated that infection with PiCV is a crucial step in the development of young pigeon disease syndrome; however, additional factors like stress or secondary infections are needed for complete development of the clinical symptoms (Raue et al., 2005). Although the clinical significance of circoviruses in avian species is not clear in each case, the distinct tropism of the circoviruses to organs of the immune system implies an involvement in immunosuppressive diseases. In the case of the starlings investigated in this study, the StCV genome was detected in the spleen of the birds, which might indicate an involvement in the epizootic disease observed among the wild birds. As the virus was also subsequently detected in apparently healthy birds, a causative role of StCV in this outbreak is questionable; it has to be taken into consideration, however, that all of these birds originated from the same population and that they could have been caught in an early stage of infection. Further studies with larger sample sizes of diseased and apparently healthy birds from different geographical regions will be necessary to assess the clinical significance of the StCV infection.

The novel circovirus was initially detected using a degenerate primer PCR, which was previously shown to be capable of detecting the GoCV, PiCV, CaCV and DuCV genomes (Todd et al., 2001a, b; Hattermann et al., 2003). In the latter cases, the remaining part of the circular genome had been amplified by inverse PCR; this approach, however, could not be successfully applied in the case of the StCV genome. A low amount or poor quality of viral DNA in the spleen samples investigated here or the existence of strong secondary structures within the DNA – such as the hairpin structure found within ORF-V2 (Fig. 2d) – may inhibit elongation during PCR. Using the technique of RCA, we were able to amplify and clone the whole StCV genome from the sample. RCA had been used previously for the successful detection of novel papillomaviruses (Rector et al., 2004a, 2005) and anelloviruses (Niel et al., 2005); however, in principle, all circular DNA virus genomes could be amplified by this technique. Moreover, a major advantage of the technique is that the amplification is independent of DNA sequence, making it possible to detect viruses with very low sequence similarities to known viruses.

The proofreading activity of the bacteriophage φ29 DNA polymerase used in RCA should guarantee a low mutation rate during amplification. For this enzyme, error rates between $10^{-5}$ and $10^{-6}$ have been assessed in several biochemical assays (Esteban et al., 1993). Using the φ29 DNA polymerase in a technique for amplification of human genome sequences, an error rate of $9.5 \times 10^{-6}$ was estimated by direct sequencing of 500 000 bp (Paez et al., 2004). In this study, two mutations were detected by comparison of sequences with approximately 2000 bp in length, which
Alignment of the StCV genome sequence revealed similarities of up to 67.1% to other circovirus genomes, indicating that StCV represents a novel species within the genus Circovirus. The genome shows typical features that are common to all circoviruses: ORFs encoding the Rep and capsid proteins as well as a small non-coding region containing a stem-loop structure with a nonamer consensus sequence (Todd et al., 2001b; Phenix et al., 2001). Additional ORFs are not conserved; in the case of StCV, only one additional ORF with a coding capacity of more than 100 aa was found and designated ORF-V2. The function of ORF-V2 is not clear as no significant similarities to other functional protein sequences were found. The stem-loop structure in StCV has a length of 17 nt; only that of PICV is longer (20 nt, Mankertz et al., 2000). This structure is thought to represent the origin of rolling-circle replication of the viral genome (Phenix et al., 2001), but the significance of the length of the stem–loop is not known. Although the highest degree of sequence similarity of StCV is found with CaCV, the length of the CaCV stem–loop is only 8 nt (Phenix et al., 2001).

The phylogenetic analysis revealed the closest relationship of StCV to CaCV. The natural hosts of these viruses – starling and canary – are also closely related as they are both members of the order Passeriformes. To analyse the relationships of the circovirus hosts in more detail, an alignment of a fragment of their genome sequence was performed. The evolutionarily conserved cytochrome b sequence was chosen due to its availability for a large number of species and because of its previous successful use for a variety of phylogenetic analyses (Kocher et al., 1989) including those of bird species (Griffiths et al., 2004) and those of virus–host relationships (Hughes & Friedman, 2000; Dekonenko et al., 2003). Interestingly, the phylogenetic tree established for the cytochrome b gene fragment of the hosts resembles that of the circoviruses. This might indicate a host-specific evolution of the circoviruses in which no virus transmission took place between distantly related hosts. However, slight differences in the branching of the trees as well as major differences in the distances between the sequences in both trees have been observed. The latter finding confirms that the cytochrome b gene is highly conserved (Kocher et al., 1989); however, a high degree of variability in the circovirus sequences is also evident. It is well known that some of the circoviruses, e.g. BFDV, show a high degree of sequence heterogeneity with a tendency to develop species-specific genotypes (Raue et al., 2004; de Kloet & de Kloet, 2004; Heath et al., 2004). Further analysis of circovirus field-origin sequences as well as experimental infection trials are needed to assess the routes of circovirus transmission and to reconstruct the evolution of these viruses. The application of the RCA technique may be helpful in these investigations.

ACKNOWLEDGEMENTS

The authors wish to thank Michael Hofreiter (Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany) for helpful discussions about phylogenetic analyses. Also, we wish to thank Erhard F. Kaleta (Justus Liebig University, Giessen, Germany) and the Central Veterinary Laboratory of Algete (Spain) for collaboration. We acknowledge the City Council of Huesca (Aragon, Spain) for providing the facilities for picking up the birds and for data collection. The study was financed in part by the Government of Aragon (project ‘Health status of game wildlife’).

REFERENCES

de Kloet, E. & de Kloet, S. R. (2004). Analysis of the beak and feather disease viral genome indicates the existence of several genotypes which have a complex psittacine host specificity. Arch Virol 149, 2393–2412.


