Whole-genome characterization of a novel polyomavirus detected in fatally diseased canary birds

Mohammad Yahya Halami,1 Gerry M. Dorrestein,2 Peter Couteel,3 Gerald Heckel,4,5 Hermann Müller1 and Reimar Johne6

1Institute for Virology, Faculty of Veterinary Medicine, University of Leipzig, An den Tierkliniken 29, 04103 Leipzig, Germany
2Dutch Research Institute for Birds and Exotic Animals (NOIVBD), Veldhoven, The Netherlands
3Veterinary Praxis Trigenio, Nijlen, Belgium
4Computational and Molecular Population Genetics (CMPG), Institute of Ecology and Evolution, University of Bern, Bern, Switzerland
5Swiss Institute of Bioinformatics, Genopode, Lausanne, Switzerland
6Federal Institute for Risk Assessment, Diedersdorfer Weg 1, 12277 Berlin, Germany

Correspondence
Mohammad Yahya Halami
halami@vetmed.uni-leipzig.de

Received 12 May 2010
Accepted 23 August 2010

Polymaviruses of birds are aetiological agents of acute inflammatory diseases in non-immunocompromised hosts, which is in contrast to mammalian polymaviruses. VP4, an additional structural protein encoded by the viral genomes of the known avian polymaviruses, has been suggested to contribute to pathogenicity through loss of cells following induction of apoptosis. Four distinct bird polymaviruses have been identified so far, which infect crows, finches, geese and parrots. Using broad-spectrum PCR, a novel polymavirus, tentatively designated canary polyomavirus (CaPyV), was detected in diseased canary birds (Serinus canaria) that died at an age of about 40 days. Intranuclear inclusion bodies were found in the liver, spleen and kidneys. The entire viral genome was amplified from a tissue sample using rolling-circle amplification. Phylogenetic analysis of the genome sequence indicated a close relationship between CaPyV and other avian polymaviruses. Remarkably, an ORF encoding VP4 could not be identified in the CaPyV genome. Therefore, the mechanism of pathogenicity of CaPyV may be different from that of the other avian polymaviruses.

INTRODUCTION

Polyomaviruses of birds are agents of inflammatory disease in birds (Johne & Müller, 2007). Acute clinical diseases with high mortality rates have been recorded after natural and experimental infection with avian polymavirus (APV) in budgerigars (Krautwald et al., 1989) and with goose haemorrhagic polymavirus (GHPV) in geese (Guérin et al., 2000). The major pathological findings were hepatitis, splenomegaly and hydropneumonia during APV infection and haemorrhagic nephritis during GHPV infection. In the case of finch polymavirus (FPyV), the co-incidence of virus detection and an acute disease in bullfinches suggest an aetiological role for this virus also (Johne et al., 2006). In addition, chronic disease of the skin and feathers has been observed after infection with FPyV and APV (Krautwald et al., 1989; Wittig et al., 2007). In contrast, mammalian polymaviruses generally cause innocuous infections in non-immunocompromised hosts. Tumour induction after inoculation of laboratory rodents is a well-known feature of these viruses (Sariyer et al., 2004).

The family Polyomaviridae comprises a variety of different polymaviruses which infect humans, monkeys, cattle, rabbits, rats, mice, hamsters and various bird species mainly in a host-specific manner (Hou et al., 2005; Johne et al., 2005, 2006; Allander et al., 2007; Gaynor et al., 2007). Among these, APV, GHPV, FPyV and crow polymavirus (CPyV) represent the known polymaviruses of birds. Phylogenetic, structural and biological data support the classification of bird polymaviruses in a separate group within the family Polyomaviridae (Stoll et al., 1993; Johne & Müller, 1998; Johne & Müller, 2003). Very recently, two additional polymaviruses were detected in Bornean and Sumatran orangutans (Groenewoud et al., 2010). Many of these polymaviruses have been detected by molecular
techniques using transcriptome subtraction (Feng et al., 2008), broad-spectrum PCR (Groenewoud et al., 2010; Johne et al., 2005) or rolling-circle amplification (RCA) of the whole genome (Johne et al., 2006, 2009).

Polyomaviruses are characterized by having a non-enveloped icosahedral capsid and a circular dsDNA genome that is approximately 5000 bp in size (Hou et al., 2005). Early and late genes are transcribed bi-directionally starting from a short non-coding regulatory region. The early genes encode the tumour (T) antigens which participate in viral genome replication and transformation of cells. The late genes encode the major structural protein VP1 and the minor structural proteins VP2 and VP3. The latter two proteins are expressed from the same ORF by using different initiation codons. In addition, some of the primate and human polyomaviruses encode the so-called agnoprotein. This non-structural protein is dispensable for propagation in cell culture (Shenk et al., 1976). Recently, an additional protein, consisting of the 125 C-terminal amino acid residues of VP3 and designated VP4, has been identified in cells infected with the monkey polyomavirus SV40 (Daniels et al., 2007).

In the genomes of bird polyomaviruses, an additional ORF is present upstream of the VP2-encoding region that encodes a protein designated VP4 in APV or a homologous protein in the other bird polyomaviruses (Luo et al., 1995; Johne & Müller, 2007). In spite of their having the same designation, VP4 of SV40 and VP4 of APV are distinctly different proteins that are encoded by completely separated ORFs. VP4 of APV is regularly present in purified viral particles and interactions with VP1 as well as with dsDNA have been identified (Johne & Müller, 2001). In addition, induction of apoptosis has been linked to expression of VP4 (Johne et al., 2000). APV mutants with deletions in the VP4-encoding region show deficiencies in virus release and have reduced infectivity (Johne et al., 2007). Due to these properties, VP4 and its homologues have been proposed to represent pathogenicity factors, which may contribute to the induction of acute disease in bird polyomaviruses (Johne & Müller, 2007).

In this study investigations on the aetiological agent of a fatal disease of canary birds (Serinus canaria) revealed the presence of a novel polyomavirus. Whole-genome analysis of the virus was performed to characterize the virus and to enable development of diagnostic tests for its detection. The genome sequence should give more insights into the phylogenetic relationships among the polyomaviruses and enable the identification of possible mechanisms of pathogenicity for the bird polyomaviruses.

RESULTS

Clinical findings and post-mortem examination

In an aviary of 50 breeding pairs of colour canaries (S. canaria), only a low number of eggs was recorded within the first breeding period. In the second breeding round, the number of eggs was within normal limits, but the hatchability was below expectation. In hatched canaries mortality occurred at an age of about 40 days and about 50 % of the nestlings died. Only two specific colour types were affected: agate yellow opal and agate white opal. The affected birds were seen to be ill for 1 day. At necropsy, the most prominent findings were subcutaneous haemorrhages, hepato- and splenomegaly (Fig. 1a) and an enlarged bursa of Fabricius (Fig. 1b). In the liver, extensive centrolobular degeneration with haemorrhages was seen and extramedullar haemopoiesis was present in the sinuosids (Fig. 1c). In these areas enlarged ‘empty’ nuclei were seen. These resembled typical polyomavirus-like intranuclear inclusion bodies (Fig. 1d). The spleen showed depletion and polyomavirus-like intranuclear inclusion bodies (not shown). In addition, polyomavirus-like intranuclear inclusion bodies were incidentally found in the epithelium of some renal glomeruli. No intranuclear or intracytoplasmic inclusion bodies were seen in the bursa of Fabricius. In addition to these findings, macrophages or monocytes containing typically shaped Atoxoplasma (Isospora serini) trophozoites were detected in the lung and the spleen.

Detection of a novel polyomavirus in samples of diseased birds

Paraffin-embedded samples containing multiple organs (liver, kidney, spleen, lungs, heart and bursa of Fabricius) of the canary birds were tested for the presence of polyomavirus DNA using a nested broad-spectrum PCR.

![Fig. 1. Post-mortem findings in canary birds that died at an age of about 40 days. (a, b) Necropsy shows enlarged liver and spleen (a) and an enlarged bursa of Fabricius (b, marked by an arrow). (c, d) Histopathological examination of a liver sample identifies haemorrhages (c) and intranuclear inclusion bodies (d), as indicated by arrows.](http://vir.sgmjournals.org)
DNA isolated from a GHPV-positive kidney served as a positive control. After electrophoresis, a PCR product with the expected size of about 270 bp was detected in the samples (Fig. 2a). Sequence analysis of this product, located in the VP1 genomic region, revealed identities \(\leq 71\%\) with the genome sequences of polyomaviruses (not shown). The novel virus was tentatively designated canary polyomavirus (CaPyV).

**Amplification of the CaPyV genome**

In order to amplify the circular viral genome, DNA from the liver sample was subjected to multiply primed RCA. After digestion of the RCA product with \(Pst\)I, multiple bands were visible. Only a very slowly migrating band, corresponding to undigested DNA, was visible after \(Eco\)RI treatment (Fig. 2b). A single band of approximately 5 kbp, generated by digestion with \(Xho\)I and considered to represent a linearized polyomavirus genome, was subsequently cloned and sequenced. Analysis of the sequence revealed that this fragment contained the sequence of the PCR product mentioned above, thus indicating that – according to the protocol applied – the complete CaPyV genome had been cloned.

**Analysis of the genome sequence**

The genome of CaPyV has a length of 5421 bp. ORFs encoding proteins with homologies to VP1, VP2 and VP3, as well as large and small T antigens are found in positions typical of polyomavirus genomes (Fig. 3). No ORF encoding a middle T antigen is present. An intensive search, which also took possible splicing sites into account, did not identify an ORF encoding a protein with homology to APV VP4 or to the primate polyomavirus agnoproteins. Despite the presence of a relatively long region of 1081 bp between the initiation codons for VP2 and the T antigens, only four ORFs with coding capacities of 43–65 aa in the sense direction and one ORF encoding 111 aa in the antisense direction could be found within this region. None of these ORFs showed homology with known proteins by BLAST search.

Comparison of the CaPyV genome sequence with those of 17 other polyomaviruses revealed that CaPyV is most closely related to FPv and AVP, with 51.3 and 51.2% nucleotide sequence identity, respectively. A phylogenetic tree reconstructed on the basis of complete polyomavirus genome sequences shows that all avian viruses cluster in a separate branch with high bootstrap support (Fig. 4a). A separate avian cluster including CaPyV is also present in the phylogenetic tree based on the deduced amino acid sequences for large T antigen (Fig. 4b). The shorter sequences of VP2 indicated the closer relationships among avian sequences as well, but with relatively low bootstrap support (Fig. 4d). For VP1, there was no evidence of a monophyly of all avian polyomaviruses including CaPyV (Fig. 4d). A closer relationship between GHPV and CPyV as well as between AVP and FPv was supported by analogy with the results of the analysis of the other genome regions. Generally, VP1 provided only poor resolution of the basal topology of the phylogenetic tree and bootstrap support was generally lower than for the trees based on the longer sequences of the other genome regions.

**Fig. 2.** Detection of a novel polyomavirus in a liver sample of a diseased canary bird. (a) Analysis of secondary products of the nested broad-spectrum PCR from a negative control (−), the liver sample (canary) and a positive control (+). The position of the specific band is marked by an arrow. (b) Analysis of the RCA product of the liver sample after digestion with \(Pst\)I, \(Eco\)RI or \(Xho\)I. The arrow indicates the genome-sized \(Xho\)I product, which was subsequently cloned. M, Molecular mass markers.

**Fig. 3.** Genome map of CaPyV. ORFs which putatively encode large T antigen (large T-Ag), small T antigen (small T-Ag), VP1, VP2 and VP3 are marked by arrows, with nucleotide positions indicated. The coding capacities of the ORFs are shown in parentheses in terms of number of amino acids.
A more detailed analysis of the amino acid sequences of CaPyV revealed highly conserved sequences in functionally important regions of the encoded proteins. The large T antigen shows a typical HPDKGG box and the LXCXE motif, which are conserved among most of the polyomaviruses. The nuclear localization signal (NLS) and the consensus sequence for protein phosphatase 2A binding is not found in the small T antigen of CaPyV. VP2 of CaPyV has an N-terminal signal sequence for myristoylation and a consensus sequence for the large T antigen in bird polyomaviruses. As for all bird polyomaviruses, the CXCX2C consensus sequence for protein phosphatase 2A binding is not found in the small T antigen of CaPyV. VP2 of CaPyV has an N-terminal signal sequence for myristoylation and an accumulation of basic amino acid residues which probably function as NLS at the C terminus; both of these are conserved among most of the polyomaviruses. Also, analysis of the amino acid sequence of VP1 did not reveal any unusual features. A separate alignment of the VP1 sequences of all bird polyomaviruses showed the relatively high amino acid sequence divergence of CaPyV compared with the other bird polyomaviruses (Fig. 5). Variation included several insertions/deletions and exchanges of amino acids in the N-terminal and C-terminal regions as well as in the highly variable outer loops of this capsid protein.

A search for functional sequences in the non-coding regions identified the sequence CCCATAAAAGGC at nt 186–195, which corresponds to the consensus sequence NCC(A/T)GNN, which usually serves as the binding sequence for the large T antigen in bird polyomaviruses. Signal sequences for polyadenylation of the early and late mRNAs are present at nt 3313–3308 and 3116–3321, respectively.

**DISCUSSION**

Mammalian polyomaviruses are known to cause subclinical persistent infections in their natural hosts which may progress to diseases after severe immunosuppression (Cole & Conzen, 2001). In contrast, APV and GHPV are the primary aetiological agents of severe diseases with high mortality rates in birds (Johne & Müller, 2007), and FPV has repeatedly been detected in acutely diseased birds.
Here we detected a novel polyomavirus in canary birds suffering from unexplained mortality at the age of about 40 days. Analysis of the viral genome revealed that the polyomaviruses of birds are genetically diverse; however, they share the ability to induce fatal disease as a common feature.

The major post-mortem findings observed in the canary birds were hepatitis and subcutaneous haemorrhages, which are reminiscent of acute APV infection in young budgerigars. Also, the presence of intranuclear inclusion bodies is a feature common to both cases. Only recently, such inclusion bodies were observed in a fatal disease outbreak in canary birds that were also shown to have a concurrent poxvirus infection (Shivaprasad et al., 2009).

Although it had been speculated that a polyomavirus was present in these birds, attempts to detect APV-specific DNA sequences remained without success. It may be speculated that CaPyV, which cannot be detected by APV-specific PCRs, caused the observed inclusion bodies. Therefore, application of broadly reactive diagnostic tests is necessary in order to identify genetically diverse polyomaviruses of birds. Broad-spectrum PCRs (Völter et al., 1998; Johne et al., 2006) and multiply primed RCA (Johne et al., 2009) have been shown here and in previous studies to be suitable for this purpose.

The coincidence of the clinical picture with the presence of CaPyV, as well as the similarity of the post-mortem findings with those observed in APV infection, may indicate an aetiological role for CaPyV in the observed disease. However, from the available data, it cannot be ruled out totally that other factors contributed to disease development or that CaPyV was only present accidentally in the diseased birds. Infection with *Atoxoplasma*, also detected in the birds investigated here, is very common in canaries ranging in age from 2 to 9 months and may cause disease in these birds (Dorrestein, 2009). At necropsy, an enlarged and sometimes spotted liver (with necrosis in the acute phase) may be seen, along with a huge, dark-red-coloured spleen and an oedematous duodenum with vascularization. However, subcutaneous haemorrhages and an enlarged bursa of Fabricius are not commonly seen in *Atoxoplasma* infections in canaries. In addition, the extensive centrolobular degeneration with haemorrhages in the liver and the polyomavirus-like intranuclear inclusion bodies are not found in atoxoplasmosis. In order to assess the pathogenicity of CaPyV, experimental infections, similar to those described in the case of APV (Krautwald et al., 1989) and GHPV (Guérin et al., 2000), have to be performed when the virus can be cultivated from non-formalin-fixed tissue; unfortunately, this is not available at present.

The genetic characterization of the CaPyV genome indicated a close relationship with the other polyomaviruses of birds. In most of the phylogenetic trees, the bird polyomaviruses form a separate branch indicating the separate evolution of mammalian and avian polyomaviruses. This is most obvious by analysis of the amino acid sequences of the large T antigen, a protein which is mainly involved in tumour induction by the mammalian polyomaviruses. In contrast, tumour induction is not a common feature of the bird polyomaviruses (Johne &
Müller, 2007), which may be explained by the different structures (and functions) of their large T antigens. Sequences of the structural proteins VP1 and VP2 provided only limited or no further phylogenetic support for a distinct evolution of avian versus mammalian polyomaviruses. This may be explained by the shorter sequence lengths of these proteins and by the distinct function of VP1 as the major capsid protein, which commonly has highly variable regions as a result of immunological pressure. In general, however, the data of the CaPyV genome sequence analysis support the classification of polyomaviruses from birds and mammals into two separate taxonomic groups, as suggested previously (Stoll et al., 1993; Pérez-Losada et al., 2006; Johne & Müller, 2007).

Despite the obvious similarities between CaPyV and other polyomaviruses of birds, some distinguishing features also exist. This includes several insertions/deletions and exchanges at various amino acid positions in the major capsid protein VP1. As many of these positions are located in the outer loops of the protein, differences in antigenicity between CaPyV and the other bird polyomaviruses have to be expected. However, the most obvious difference between CaPyV and the other bird polyomaviruses is the absence of an ORF encoding a VP4 homologue. VP4 of APV has been shown to be essential for effective virus replication in tissue culture; several functions in capsid assembly, induction of apoptosis and regulation of viral gene expression have also been demonstrated for this protein (Johne et al., 2000; Johne & Müller, 2001; Li et al., 2009). VP4 and its homologues have been proposed to be generally involved in the pathogenesis of disease due to infections with the bird polyomaviruses (Johne & Müller, 2007). The presence of CaPyV in acutely diseased birds in the absence of a VP4-like protein suggests a different mechanism of pathogenicity for CaPyV. Further laboratory and field studies are needed to elucidate the pathogenesis of CaPyV-associated disease. The genome sequence presented here may be used for the identification of candidate factors of pathogenicity and also serve as a template for the development of diagnostic tools for the detection and survey of CaPyV infections.

**METHODS**

**Samples.** Samples of heart, liver, spleen, lungs, kidneys and bursa of Fabricius of two-colour canaries (S. canaria) that had died at about 40 days of age in an aviary in the Netherlands in 2007 were collected and fixed in 10% buffered formalin. The fixed samples were used either directly for DNA isolation or embedded in paraffin and deoxygenated followed by histological processing.

Nested broad-spectrum PCRs. DNA was isolated from paraffin-embedded and formalin-fixed tissue samples using a DNeasy Tissue kit (Qiagen) after thorough washing in PBS (pH 7.2) with frequent changes. A nested broad-spectrum PCR for the detection of polyomavirus-specific DNA was performed as described previously (Johne et al., 2005) by using primers VP1-1f and VP1-1r, followed by nested PCR with primers VP1-2f and VP1-2r. Secondary PCR products of approximately 270 bp were sequenced directly using the PCR primers.

**Multiple primed RCA.** DNA isolated from the samples was directly amplified by multiple primed RCA using a TempliPhi 100 Amplification kit (Amersham Biosciences) as described previously (Johne et al., 2006). For restriction enzyme analysis of the RCA product, 2 μl aliquots were digested with EcoRI, PstI or XhoI (New England Biolabs) and subjected to gel electrophoresis on ethidium bromide-stained 1% agarose gels. For cloning, a total of 30 μl of the RCA product was digested with XhoI. The resulting 5 kb product was ligated with XhoI-digested pBluescript II SKI (+) vector (Stratagene). Both strands of the insert were sequenced by the primer walking method.

**Sequence analysis.** The genome sequence of CaPyV was reassembled from the sequence fragments using the SeqBuilder module of the DNAStar software package (Lasergene) and deposited in GenBank with the accession number GU345044. For sequence alignment and phylogenetic analyses, the genome sequences of 17 polyomaviruses were downloaded from GenBank (GenBank accession numbers are shown in Supplementary Table S1, available in JGV Online). ORFs were identified and amino acid sequences were deduced from the nucleotide sequences using the SeqBuilder module. Sequence alignments and reconstruction of phylogenetic trees were performed using the MEGALIGN module of DNASTAR. The CLUSTAL W method was used with the IUB (nucleotide) or Gonnet 250 (amino acid) residue-weight tables (Thompson et al., 1994) in alignments. Phylogenetic trees were reconstructed from the whole genomes and different segments using the neighbour-joining algorithms implemented in DNASTAR with complete deletion of positions containing insertion/deletion polymorphisms. Robustness of phylogenetic trees was tested by performing 1000 bootstrap replicates with 111 random seeds.

**ACKNOWLEDGEMENTS**

This work was funded by a grant from the Deutsche Forschungsgemeinschaft (JO 369/3-2).

**REFERENCES**


