Novel polyomaviruses in South American bats and their relationship to other members of the family Polyomaviridae

Zahra Fagrouch,¹ Roya Sarwari,¹ Anne Lavergne,² Marguerite Delaval,³ Benoit de Thoisy,² Vincent Lacoste² and Ernst J. Verschoor¹

¹Department of Virology, Biomedical Primate Research Centre (BPRC), Lange Kleiweg 161, 2288GJ Rijswijk, The Netherlands
²Laboratoire des Interactions Virus–Hôtes, Institut Pasteur de la Guyane, 23 avenue Pasteur, BP6010, 97306 Cayenne Cedex, French Guiana
³ONF Guyane, Service développement Sylvétude, Réserve Montabo, BP7002, 97307 Cayenne Cedex, French Guiana

Bats are the natural reservoir of a variety of viruses, including a polyomavirus (PyV) from a North American brown bat. We investigated 163 spleen samples from 22 bat species from French Guiana for the presence of PyVs. In total, we detected 25 PyV-positive animals belonging to nine different bat species. Phylogenetic analysis was performed on the genomes of eight representative PyVs, and showed that the bat PyVs form three distinct lineages within the genus Orthopolyomavirus and are genetically different from the previously described North American bat virus. Interestingly, two lineages cluster with PyVs found in chimpanzees, orangutans and gorillas. In addition, one group of bat PyVs is genetically related to the human Merkel cell polyomavirus.

Polyomaviruses (PyVs) are small DNA viruses with a genome size of approximately 5 kb that infect mammals and birds. In birds, PyVs cause acute, fatal infections, but in mammals these viruses primarily establish lifelong infections that can develop into fatal infections in immunosuppressed hosts (Krumholz et al., 2009).

Recently, the family Polyomaviridae has undergone several taxonomic revisions (Johne et al., 2011). The family now contains three genera. The genus Avipolyomavirus comprises all avian PyVs, and its distinction from the mammalian viruses is based on biological differences (host range, cell tropism, pathogenicity), as well as genomic differences.

The genera Orthopolyomavirus and Wukipolyomavirus are formed by PyVs from mammals, and the differentiation between these genera is based on nucleotide divergence. The genus Wukipolyomavirus is small and contains only four human viruses: WUPyV, KIPyV, HPyV6 and HPyV7, while the genus Orthopolyomavirus consists of a variety of viruses from primates, including humans, cattle, rodents, sea lions and a bat PyV.

Worldwide, >1200 species of bat (order Chiroptera) have been described, divided into two suborders, Megachiroptera and Microchiroptera (Tudge, 2000). They represent 26% of all presently known mammalian species, making it the second-largest group of mammals after the rodents (Wilson & Reeder, 2005). Bats are natural reservoirs of a large collection of zoonotic viruses, such as lyssaviruses, henipaviruses, the SARS coronavirus and ebolavirus (Calisher et al., 2006). More recently, bat species have been implicated as carriers of adenoviruses (Drexler et al., 2011; Jánoska et al., 2010; Li et al., 2010), herpesviruses (Jánoska et al., 2010), astroviruses (Drexler et al., 2011), picornaviruses (Lau et al., 2011) and Marburg virus (Kuzmin et al., 2010).

Misra et al. (2009) described a PyV in the North American bat species Myotis lucifugus. This finding led us to hypothesize that the primate PyVs may have originated from ancestral bat or rodent PyVs (Groenewoud et al., 2010). To support our hypothesis, we initiated a survey for PyV infection in bats. All bats examined in this study were collected as part of a field study in French Guiana, South America, for the incidence of rabies virus infection.

DNA was isolated from spleen tissue of 163 individuals belonging to 22 different bat species (Table 1). Species identification was done by sequence analysis of part of the cytochrome oxidase subunit 1 gene (Clare et al., 2007; Ivanova et al., 2006). Ten species were each represented by a single sample, but others, such as Seba’s short-tailed bat (Carollia perspicillata) and the flat-faceted fruit-eating bat (Artibeus planirostris), were represented by multiple spleen
samples. All DNA samples were screened by using two pan-
PyV PCR assays (Table S1, available in JGV Online). The
first assay has been widely used for the detection of viruses
belonging to the genus Orthopolyomavirus (Groenewoud
et al., 2010; Johne et al., 2005; Johne & Müller, 2007;
Leendertz et al., 2011; Misra et al., 2009; Verschoor et al.,
2008). The second assay targets a region of the VP1 gene
that partially overlaps with the target sequence of the first
assay. PyV-positive bats were detected in nine different bat
species (41 %; 25 PyV-positive animals in total). PyVs from
A. planirostris and from the common vampire bat (Des-
modus rotundus) were detected by both PCR assays, while
PyV from Pallas’s mastiff bat (Molossus molossus) could
only be detected by the first assay, and those from six species
(C. persicillata, Eptesicus furinalis, Glossopha ga soricina,
Platyrhinus brachycephalus, Pteronot us parrnelli and Sturnira
lilium) were detected exclusively with the second assay. This
finding implied substantial sequence variation of the bat
viruses that was confirmed by sequencing of the PCR
amplicons. A 119 bp VP1 gene segment, formed by the
overlap of the amplicons from the two assays, was used to
construct a phylogenetic tree (Figs S1 and S2). Several
well-supported sequence clusters can be distinguished; the
biggest cluster is formed by ten C. persicillata sequences,
in addition to VP1 sequences from A. planirostris R104
and R79, and G. soricina R95 (97 % bootstrap support).
Within this group, the C. persicillata sequences form a sub-
cluster (92 % bootstrap) and cluster with A. planirostris R104
(91 % bootstrap). A second cluster is formed by three A.
planirostris viruses, R94, E1030 and A504 (82 % boot-
strap), while the remaining A. planirostris sequence (A1055)
is part of a third group of six viruses identified from five
different host species (87 % bootstrap support). The residual
sequences (P. parrnelli R266, C. persicillata R107 and D.
rotundus AT7) do not cluster in our analysis. The tree was
used to select viruses representing all major clusters and
lineages for genome amplification and sequencing. The
selected viruses are indicated in bold italics in Fig. S1. For
eight viruses, the nearly full-length genome was amplified,
especially as described previously (Verschoor et al., 2008),
and complete genome sequences were obtained upon com-
bination with the diagnostic PCR sequences. Due to the lack
of material for C. persicillata (R107) or for unknown
reasons (A. planirostris R79 and G. soricina R95), no genome
sequence could be amplified. Some caution is necessary due
to the fact that our genome amplification direct from tissue
DNA may have been prone to Taq polymerase-induced chimerization, resulting in the amplification of artefactual
viruses (Liu et al., 2010).

The molecular characteristics of the novel PyVs are sum-
marized in Table 2. All genomes have a typical PyV
organization, with an early region encoding the small T-
and large T-antigens (t-Ag and T-Ag), and a late region
encoding the structural proteins VP1 and VP2/3. Viruses
A1055, B0454 and B1130 are also capable of encoding an

Table 1. Bat species tested and results of diagnostic VP1 PCR assays

<table>
<thead>
<tr>
<th>Species</th>
<th>English name</th>
<th>No. of samples</th>
<th>No. of positive samples*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artibeus cinereus</td>
<td>Gervais’s fruit-eating bat</td>
<td>5</td>
<td>0/0</td>
</tr>
<tr>
<td>Artibeus gnomus</td>
<td>Gnome fruit-eating bat</td>
<td>1</td>
<td>0/0</td>
</tr>
<tr>
<td>Artibeus obscurus</td>
<td>Dark fruit-eating bat</td>
<td>2</td>
<td>0/0</td>
</tr>
<tr>
<td>Artibeus planirostris</td>
<td>Flat-faced fruit-eating bat</td>
<td>39</td>
<td>5/6</td>
</tr>
<tr>
<td>Carollia persicillata</td>
<td>Seba’s short-tailed bat</td>
<td>60</td>
<td>0/11</td>
</tr>
<tr>
<td>Cormura brevirostris</td>
<td>Chestnut sac-winged bat</td>
<td>2</td>
<td>0/0</td>
</tr>
<tr>
<td>Desmodus rotundus</td>
<td>Common vampire bat</td>
<td>10</td>
<td>1/1</td>
</tr>
<tr>
<td>Eptesicus furinalis</td>
<td>Argentine brown bat</td>
<td>1</td>
<td>0/1</td>
</tr>
<tr>
<td>Eumops auripendulus</td>
<td>Black bonneted bat</td>
<td>1</td>
<td>0/0</td>
</tr>
<tr>
<td>Lasiusus blossevillii</td>
<td>Western red bat</td>
<td>1</td>
<td>0/0</td>
</tr>
<tr>
<td>Glossopha ga soricina</td>
<td>Pallas’s long-tongued bat</td>
<td>1</td>
<td>0/1</td>
</tr>
<tr>
<td>Molossus molossus</td>
<td>Pallas’s mastiff bat</td>
<td>7</td>
<td>1/0</td>
</tr>
<tr>
<td>Phylodroma stenops</td>
<td>Pale-faced bat</td>
<td>1</td>
<td>0/0</td>
</tr>
<tr>
<td>Phylllostomus discolor</td>
<td>Pale spear-nosed bat</td>
<td>3</td>
<td>0/0</td>
</tr>
<tr>
<td>Platyrhinus brachycephalus</td>
<td>Short-headed broad-nosed bat</td>
<td>2</td>
<td>0/1</td>
</tr>
<tr>
<td>Platyrhinus helleri</td>
<td>Heller’s broad-nosed bat</td>
<td>1</td>
<td>0/0</td>
</tr>
<tr>
<td>Pteronotus parrnelli</td>
<td>Parnell’s mustached bat</td>
<td>1</td>
<td>0/1</td>
</tr>
<tr>
<td>Rhinophylla pumilio</td>
<td>Dwarf little fruit bat</td>
<td>1</td>
<td>0/0</td>
</tr>
<tr>
<td>Saccopteryx bilineata</td>
<td>Greater sac-winged bat</td>
<td>1</td>
<td>0/0</td>
</tr>
<tr>
<td>Sturnira lilium</td>
<td>Little yellow-shouldered bat</td>
<td>11</td>
<td>0/2</td>
</tr>
<tr>
<td>Sturnira tildae</td>
<td>Tilda’s yellow-shouldered bat</td>
<td>7</td>
<td>0/0</td>
</tr>
<tr>
<td>Uroderma bilabatum</td>
<td>Tent-making bat</td>
<td>5</td>
<td>0/0</td>
</tr>
</tbody>
</table>

*Assay 1 (Johne et al., 2005)/assay 2 (this manuscript).
agnoprotein, a characteristic that they share with avian PyVs and several primate viruses (Deuzing et al., 2010; Johne et al., 2011; Khalili et al., 2005; Verschoor et al., 2008). Remarkably, the genome of A1055 contains two potential agnogenes in two different ORFs. In contrast, the bat viruses from A. planirostris R104 and C. perspicillata C1109 share the lack of an agnogene with the human Merkel cell polyomavirus (MCPyV), a gorilla virus (GggPyV1) and two other chimpanzee PyVs (ChPyV1a and -2a). Remarkably, the early region of A1055 contains a single, long ORF encoding a t-Ag of 849 aa, which shares its 3’ terminus with the predicted T-Ag. A similar, unusually long t-Ag ORF was also described for the chimpanzee PyV isolate Azzie (Deuzing et al., 2010). We performed splice analysis of the early-region mRNA with a neural network splice-site prediction program on the NetGene2 server (Brunak et al., 1991; Hebsgaard et al., 1996). Computational analysis of the translated spliced mRNA indicated that virus A1055 probably employs a variant splicing of the t-Ag. To acquire a t-Ag with a high similarity to other bat PyVs, the early transcript is spliced at a strong donor splice site within the coding sequence (position 4479), instead of downstream of the termination codon. Two predicted acceptor splice sites, one at a position just upstream of the mRNA stop codon (position 4413) and the second at position 3642, may be used to complete the t-Ag mRNA. This results in a small T-antigen of 183 or 182 aa in length, respectively. Such a splicing pattern for the t-Ag mRNA has been described for the murine and hamster PyVs (MPyV and HaPyV, respectively) and for MCPyV (Shuda et al., 2008). However, the t-Ag ORF of MCPyV is situated entirely in the first exon, in contrast to the t-Ag ORFs of A1055 and the rodent viruses, where the translational stop codon is located in the second exon. While being present in murine and hamster PyVs, no evidence for a middle T-antigen spliced mRNA was found in our analysis of A1055 (Delmas et al., 1985; Hunter et al., 1979).

To investigate the evolutionary relationships of the new bat viruses with other mammalian PyVs in more detail, a phylogenetic tree was constructed on the basis of the aligned complete genomes (Fig. 1). Additionally, we performed phylogenetic analysis separately on the early T-Ag-encoding region, and the late region genes, VP1 and VP2 (Figs S3 and S4). The overall topology of the evolutionary trees was comparable, but some differences could be observed. The newly characterized viruses all belong to the genus Orthopolyomavirus, and in all analyses they separated into three distinct, well-supported genetic clusters (A, B and C). In the analysis using complete genomes, the group A viruses were part of a cluster with PyVs from a variety of species, including sea lions, cattle, monkeys, rodents and the North American bat. Further analyses using individual genes showed a similar branching pattern, but only smaller (sub)clusters were supported by bootstrap values of >70 %. Of particular interest was the position of the cluster formed by JCPyV, BKPyV, SA12 and SV40-Ri257. Depending on the analysis performed, this group was either linked to the group A viruses (85 % bootstrap) (T-Ag analysis, Fig. S3) or could be found as a separate lineage in variable positions on the tree (Figs 1 and S4). In Fig. 1, the B/C bat viruses are strongly clustered (88 % bootstrap) with great ape PyVs from orangutans, chimpanzees and gorillas, and with the human trichodysplasia spinulosa-associated polyomavirus (TSPyV) and MCPyV (Deuzing et al., 2010; Feng et al., 2008; Groenewoud et al., 2010; Leendertz et al., 2011; van der Meijden et al., 2010). In the other analyses, this cluster was not supported by bootstrap analysis (VP1, Fig. S4), or the viruses were part of a larger cluster that also included the rodent viruses HaPyV and MPyV, (T-Ag; Fig. S3) or the viruses belonging to the LPyV-HPyV9 lineage (VP2; Fig. S4).

In the phylogenetic trees, the bat viruses from cluster C are related closely to MCPyV and its counterparts circulating in great apes (ChPyV1a, ChPyV 2a and GggPyV1) (Leendertz et al., 2011). A feature common to these viruses is a relatively large genome (>5300 nt) and – but this is

### Table 2. Genome features of bat PyVs

<table>
<thead>
<tr>
<th>Origin</th>
<th>Virus</th>
<th>Cluster</th>
<th>Genome size (nt)</th>
<th>ORF (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pteronotus parnellii</em></td>
<td>R266</td>
<td>A</td>
<td>5041</td>
<td>175</td>
</tr>
<tr>
<td><em>Desmodus rotundus</em></td>
<td>A17</td>
<td>A</td>
<td>5201</td>
<td>175</td>
</tr>
<tr>
<td><em>Artibeus planirostris</em></td>
<td>A504</td>
<td>A</td>
<td>5187</td>
<td>175</td>
</tr>
<tr>
<td><em>Artibeus planirostris</em></td>
<td>A1055</td>
<td>B</td>
<td>5019</td>
<td>185</td>
</tr>
<tr>
<td><em>Stenura lilium</em></td>
<td>B0454</td>
<td>B</td>
<td>5058</td>
<td>205</td>
</tr>
<tr>
<td><em>Molossus molossus</em></td>
<td>B1130</td>
<td>B</td>
<td>4903</td>
<td>191</td>
</tr>
<tr>
<td><em>Artibeus planirostris</em></td>
<td>R104</td>
<td>C</td>
<td>5371</td>
<td>184</td>
</tr>
<tr>
<td><em>Carollia perspicillata</em></td>
<td>C1109</td>
<td>C</td>
<td>5352</td>
<td>184</td>
</tr>
</tbody>
</table>

*Variant splicing of small T-antigen mRNA.
†The A1055 genome has two putative agnogenes.
not exclusive to this cluster – the lack of a clear agnoprotein ORF. Cluster B viruses have the smallest genome of the bat viruses (Table 2), and are genetically related to another chimpanzee virus (ChPyV-Bob) (Deuzing et al., 2010). Interestingly, viruses from this lineage are characterized by the presence of a putative agnogene in their genome, which discriminates them from the group A and C viruses.

The PyV from *Myotis lucifugus* was officially named bat PyV (BatPyV) by the Polyomavirus Study Group of the International Committee on Taxonomy of Viruses (ICTV) (Johne et al., 2011). According to the demarcation criteria set by the ICTV, a novel PyV species should have <81–84 % sequence identity to other PyV genomes. The viruses described here all have <81 % sequence identity, except for A1055 and B0454, which have 84 % identity to each other. We propose a revised nomenclature: BatPyV1 for the original Myotis virus, and BatPyV2, -3 and -4 for those belonging to clusters A, B and C, respectively. PyV species within a cluster are indicated by lower-case letters (Fig. 1).

In this study we investigated samples from 22 bat species, but in French Guiana alone at least 102 species of bats have been described (Simmons & Voss, 1998). From one species, *A. planirostris*, we determined the genome of three gene-clusters; BatPyV genomes have been deposited under accession numbers JQ958886–JQ958893. Bar, 0.03 substitutions per site.

**Fig. 1.** Phylogenetic tree of the full-length genomes of mammalian PyVs. Lineages leading to bat PyVs are shaded in grey. Proposed new PyV (sub)species are indicated in bold. The sequence alignment was made by using MacVector version 10.6, and the GapStreeze program (http://www.hiv.lanl.gov/content/sequence/GAPSTREEZE/gap.html) was used to remove all columns with gaps. Phylogenetic analysis was performed using the neighbour-joining method with p-distance as implemented in MEGA version 4 (Tamura et al., 2007). Bootstrap values (as percentages of 1000 resamplings) are indicated. The tree was drawn using FigTree v. 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree). GenBank accession numbers of the viruses used in the tree are shown in parentheses; BatPyV genomes have been deposited under accession numbers JQ958886–JQ958893.
raises questions regarding the natural history of the family *Polyomaviridae*. An MCPyV-like virus in great apes led Leendertz *et al.* (2011) to speculate on the zoonotic transfer resulting in MCPyV infection in humans. Our discovery of an MCPyV-like virus in bats, related somewhat more distantly to MCPyV than to the ape viruses, is in line with our hypothesis regarding the origins of PyV infections in primates, including humans (Groenewoud *et al.*, 2010). The splicing pattern of the t-Ag mRNA from virus A1055 raises intriguing questions regarding the evolutionary relationships between the rodent and bat PyVs, and adds to the complexity of the natural history of PyVs. However, because our analyses were performed using a neural network splice-site prediction program, they have to be verified using mRNA analysis of PyV-infected cells. Clearly, only a fraction of the PyV genetic variation in various mammalian species has so far been described, and multidisciplinary research connecting the genetic variation of viruses with the geographical distribution and habitat of the viral hosts is needed to solve these issues and to verify our hypothesis.

Acknowledgements

This study was supported by the European Community Research Infrastructures Program grant RII3-CT-2006-026155 ‘European Primate Network: Specialized Infrastructures and Procedures for Biological and Biomedical Research (EUPRIM-NET)’ and by the European Regional Development Fund (ERDF) in the framework of the ViRUSES programme.

References


Polyomaviruses in South American bats


