Discovery of African bat polyomaviruses and infrequent recombination in the large T antigen in the *Polyomaviridae*

Michael Carr,1,2† Gabriel Gonzalez,3† Michihito Sasaki,4 Kimihito Ito,3 Akihiro Ishii,5 Bernard M. Hang’ombe,6 Aaron S. Mweene,7 Yasuko Orba4 and Hirofumi Sawa1,4,8,*

**Abstract**

Bat species represent natural reservoirs for a number of high-consequence human pathogens. The present study investigated the diversity of polyomaviruses (PyVs) in Zambian insectivorous and fruit bat species. We describe the complete genomes from four newly proposed African bat PyV species employing the recently recommended criteria provided by the *Polyomaviridae* Study Group of the International Committee on Taxonomy of Viruses. A comprehensive phylogenetic and recombination analysis was performed to determine genetic relationships and the distribution of recombination events in PyV from mammalian and avian species. The novel species of PyV from Zambian bats segregated with members of the genera *Alphapolyomavirus* and *Betapolyomavirus*, forming monophyletic clades with bat and non-human primate PyVs. *Miniopterus schreibersii polyomavirus 1 and 2* segregated in a clade with South American bat PyV species, Old World monkey and chimpanzee PyVs and *Human polyomavirus 13* (New Jersey PyV). Interestingly, the newly described Egyptian fruit bat PyV, tentatively named *Rousettus aegyptiacus polyomavirus 1*, had the highest nucleotide sequence identity to species of PyV from Indonesian fruit bats, and *Rhinolophus hildebrandtii polyomavirus 1* was most closely related to New World monkey PyVs. The distribution of recombination events in PyV genomes was non-random: recombination boundaries existed in the intergene region between VP1 and LTAg and also at the 3’ end of VP2/3 in the structural genes, whereas infrequent recombination was present within the LTAg gene. These findings indicate that recombination within the LTAg gene has been negatively selected against during polyomaviral evolution and support the recent proposal for taxonomic classification based on LTAg to define novel PyV species.

**INTRODUCTION**

Polyomaviruses (PyVs; family *Polyomaviridae*) are small (genome size approximately 5 kb, capsid diameter 40–50 nm), non-enveloped, circular, double-stranded (ds) DNA viruses found in a wide range of mammalian, avian, fish, insect and arachnid species [1, 2]. The PyV genome is functionally divided into an early region, encoding the small T antigen (STAg) and large T antigen (LTAg) nonstructural proteins, and a late region harbouring, minimally, the genes encoding the VP1 and VP2 capsid structural proteins and, in some PyV genomes, VP3 and agnoprotein, which are expressed following DNA replication [1]. The non-coding control region (NCCR) contains the origin of replication (Ori) where bidirectional transcription occurs and elements binding viral and cellular transcription factors for regulation of both early and late gene expression [1].

Criteria for the definition of a novel PyV species have recently been delineated by the *Polyomaviridae* Study Group of the International Committee on Taxonomy of Viruses (ICTV) [3]. A newly proposed PyV species should possess: a complete dsDNA genome sequence, with typical early and late regions present on opposite strands, encoding T antigen nonstructural and capsid structural proteins, respectively; mitochondrial *cytochrome b* (*cytb*) gene typing of the host species and, finally,
an observed genetic distance from members of the most closely related PyV species of over 15% for the LTAg coding sequence. As PyVs appear to be host-species-specific, the new nomenclature includes the binomial host genus and species, and viruses are to be numbered consecutively following the chronological order of their discovery.

Despite the stability of the PyV dsDNA genome, genetic exchange occurring in PyV genomes is now thought to have led to mosaicism following recombination events [2, 4–6]. This has prompted the move from assigning species based on less than 81% whole-genome identity to use of the LTAg region based on molecular clock analyses, which showed this nonstructural region to contain the lowest amino acid sequence variation [3]. Four novel genera have been proposed: the genera Alphapolyomavirus, Betapolyomavirus and Deltapolyomavirus, which have only been identified in mammalian species, and the genus Gammapolyomavirus, which contains only avian PyV species. Recent findings, identifying PyV genomes in fish and insect species, have indicated that modern PyV lineages arose after ancient recombination events in evolutionarily distant species [2].

Bats, order Chiroptera, are natural reservoirs for a large variety of high-impact and emerging zoonotic DNA and RNA viral pathogens [7]. Previous studies have investigated PyV diversity in bat populations in North, Central and South America, Africa, Indonesia and New Zealand [5, 8–12]. Tao and colleagues presented the important finding that bat PyV lineages were paraphyletic (dispersed across phylogenies), indicative of multiple species jumps between bats and other mammalian species [5]. Furthermore, the monophyletic relationship between bat PyVs and non-human primate (NHP) PyVs is striking, with three previously described bat–NHP monophyletic groups [5].

In the present study, we have investigated the presence of PyVs in Zambian insectivorous and fruit bat species employing a broad-spectrum PCR approach. We report the complete genomes of four new PyV species from three Old World insectivorous and fruit bat species and have performed a detailed phylogenetic and recombination analysis with all known PyVs. We have identified evidence of recombination in the newly described African PyVs and also in human, bat, NHP and other mammalian and avian species of PyV. Strikingly, the recombination events were non-random, with hot spots encompassing the structural regions encoding the viral capsid proteins with cold spots within the nonstructural gene regions. These findings may have implications for the understanding of PyV evolution and support the recent proposal to classify novel PyV species on the basis of LTAg genetic distances.

RESULTS

Identification of novel PyVs in Zambian bat species

Zambian insectivorous bat species (the common bent-wing bat *Miniopterus schreibersii* (*n*=13); horseshoe bats of the genus *Rhinolophus* [Rhinolophus simulator (*n*=1), *Rhinolophus cloequis* (*n*=1), *Rhinolophus hildebrandtii* (*n*=1)] and the giant roundleaf bat *Hipposideros gigas* (*n*=1)] and fruit-eating [*Rousettus aegyptiacus* (*n*=3)] bat species were collected at Leopard’s Hill Cave, Lusaka, Zambia in 2012 and typed by mitochondrial *cytb* gene sequencing [12]. Spleen genomic DNA was extracted and a broad-spectrum nested PyV VP1 PCR was performed [14, 15]. Six samples (four from *M. schreibersii* and one each from *Rhinolophus hildebrandtii* and *Rousettus aegyptiacus*) showed a specific PyV VP1 amplicon of approximately 250 bp. Six Zambian bat PyV full-length genomes from the three bat species were subsequently amplified by inverse PCR, cloned and sequenced completely by primer walking.

The genetic distances of the six PyV genomes from Zambian bat species with the most closely related species of PyV were calculated by comparison of the pairwise nucleotide sequence identities for the LTAg coding sequence (Table 1). Each species of Zambian bat PyV showed over 15% nucleotide sequence divergence in the LTAg coding sequence from known species of PyV and from each other (with the exception of three strains of *Miniopterus schreibersii polyomavirus 1*, thus provisionally meeting the criteria for assignment as novel species of PyV. The four novel species were tentatively named Zambian bat PyV species (*Miniopterus schreibersii polyomavirus 1, Miniopterus schreibersii polyomavirus 2, Rhinolophus hildebrandtii polyomavirus 1* and *Rousettus aegyptiacus polyomavirus 1*). Predicted open reading frames are annotated in Fig. 1. Three strains of a novel PyV species, *Miniopterus schreibersii polyomavirus 1a–c*, were identified, and in the same host a distinct PyV species was designated *Miniopterus schreibersii polyomavirus 2*, employing the host species and consecutive numbering included in the newly proposed nomenclature for PyVs [3]. *Miniopterus schreibersii polyomavirus 1* and 2 were most closely related genetically to each other; however, they were 76% identical over their entire genomes and 76% over the LTAg coding region and, thus, constitute distinct species. The six sequenced PyV genomes ranged from 4752 to 5060 bp in length, with an overall DNA G+C content of 54–59%. Genome features of the novel Zambian PyV species are described in Table 2. Each of the Zambian PyV genomes displayed the archetypal genome organization, including early and late regions, encoding nonstructural proteins, LTAg and STAg and the structural capsid proteins VP1, VP2 and VP3, respectively. These early and late regions were separated by an NCCR that is homologous to those of previously described polyomaviruses. The STAg proteins encoded by the strains of *Miniopterus schreibersii polyomavirus 1* and 2 were respectively 182 and 183 amino acid (aa) residues in length, which were slightly larger than the *Rhinolophus hildebrandtii polyomavirus 1* and *Rousettus aegyptiacus polyomavirus 1* STAg proteins, at 165 and 164 residues, respectively. The median length of 75 publicly available and the four newly described STAg sequences was 184 aa residues (range: 124–902). Each of the four novel Zambian bat PyV species STAg sequences harbours binding motifs (CXCXXC) for the protein phosphatase 2A (PP2A)
tumour suppressor, which has been shown to directly interact with the PP2A scaffolding A subunit, modifying PP2A activity and displacing regulatory B subunits from the A subunit [16].

A putative gene encoding the alternate frame of the large T ORF (ALTO), overlapping with the LTAg ORF, in the early region was identified in the Miniopterus schreibersii polyomavirus 1 and Miniopterus schreibersii polyomavirus 2 genomes and showed highest amino acid sequence identity (42 and 44%, respectively) to the ALTO protein from Human polyomavirus 13 (KF954417). On the other hand, shorter ALTO-encoding ORFs were predicted in the Rousettus aegyptiacus polyomavirus 1 and Rhinolophus hildebrandtii polyomavirus 1 genomes, of 66 and 73 amino acids, respectively (Table 2).

**Phylogenetic analysis of novel Zambian bat PyVs**

A Bayesian inference of phylogeny of the LTAg protein from recognized species of PyV with the four novel Zambian bat PyV LTAg sequences described in the present study was performed (Fig. 2). Table 1 presents the pairwise nucleotide sequence identities of the newly described Zambian bat PyVs and the most closely related PyVs identified previously from bats, NHPs and humans. The two Miniopterus schreibersii PyV genomes were 76 % identical over the LTAg coding region and clustered within the genus Alphapolyomavirus segregating in a discrete highly supported clade with PyVs from bats, NHPs and Human polyomavirus 13. Miniopterus schreibersii polyomavirus 1 and Miniopterus schreibersii polyomavirus 2 showed highest sequence identity over the LTAg to the following bat PyV species (locations/common host names are in parentheses): 67 % identity to Molossus molossus polyomavirus 1 (South and Central America/velvety free-tailed bat), 66 % identity to Artibeus planirostris polyomavirus 2 (South America/flat-faced fruit bat), 66 % identity (65 % for M. schreibersii polyomavirus 2) to Sturnira lilium polyomavirus 1 (South and Central America/little yellow-shouldered bat) and 65 % identity to Otomops martiensseni polyomavirus 1 (Africa/ Martienssen’s free-tailed bat). Interestingly, the two novel species of PyVs, identified in the Zambian common bent-wing bats, showed 65 % identity with NHP PyVs from Old World colobus monkeys (Piliocolobus rufomitratus polyomavirus 1) and 64 % identity to vervet monkey PyV (Chlorocebus pygerythrus polyomavirus 1) from Mfuwe, Zambia, 58 % identity with Pan troglodytes polyomavirus 1 and also, notably, 65 % identity (66 % for M. schreibersii polyomavirus 2) with the newly described New Jersey polyomavirus (Human polyomavirus 13) [17].

The Rhinolophus hildebrandtii and Rousettus aegyptiacus PyV LTAg amino acid sequences segregated with members of the genus Betapolyomavirus (Fig. 2). Rhinolophus hildebrandtii polyomavirus 1 and Rousettus aegyptiacus polyomavirus 1 possessed 61 % nucleotide sequence identity to each other over the LTAg coding region. The horseshoe bat Rhinolophus hildebrandtii polyomavirus 1 was most closely related genetically to New World monkey PyV species, with 65 % identity to white-fronted capuchin PyV, Cebus albifrons polyomavirus 1 and 65 % to squirrel monkey PyV, Saimiri sciureus polyomavirus 1 and also 57 % identity to Miniopterus africanus polyomavirus 1 from the African long-fingered bat. Interestingly, Rhinolophus hildebrandtii polyomavirus 1 exhibited 63 % identity over the LTAg to Mastomys natalensis polyomavirus 1 characterized from a multimammate mouse collected in Namwala in Zambia [18]. The Egyptian fruit bat Rousettus aegyptiacus polyomavirus 1 was most closely related genetically over the LTAg to Indonesian fruit bat PyVs 75 % identity to Dobsonia moluccensis polyomavirus 2 (previously named bat polyomavirus 6b, BatPyV6b) from the Moluccan naked-backed fruit bat; 66 % identity with Acerodon celebensis polyomavirus 2 (previously bat polyomavirus 6a, BatPyV6a) from the Sulawesi flying fox and 61 % identity with Dobsonia moluccensis polyomavirus 3 (previously bat polyomavirus 6c, BatPyV6c) [10]. Rousettus aegyptiacus polyomavirus 1 was also genetically related to species of PyV from New World monkeys, including Saimiri sciureus polyomavirus 1 (62 % identity) and Cebus albifrons polyomavirus 1 (61 % identity) and to the Zambian Mastomys natalensis polyomavirus 1 with 62 % identity over the LTAg coding region.

**Genome organization of the species of PyV from Zambian bats**

The NCCRs of the PyVs from Zambian insectivorous and fruit bat species contained up to eight iterations of the pentanucleotide GAGGC (or its reverse complement GCCGTC) LTAg binding motifs, including the palindromic octamer of overlapping GAGGCCCTC bidirectional LTAg binding elements, in Miniopterus schreibersii polyomavirus 1a–c (Fig. 3). Notably, the Miniopterus schreibersii polyomavirus 1a variant contained a 126 bp deletion in the NCCR which removed a predicted 3’ LTAg binding site when aligned with Miniopterus schreibersii polyomavirus 1b and 1c. The NCCRs contains downstream AT-rich regions with putative TATA boxes at approximately nucleotides 90 and 590 (Fig. 3). This indicates that both the early and late promoter may contain a TATA box for the newly described PyVs. This is in contrast with other bat PyVs (with the exception of Molossus molossus polyomavirus 1) where the late promoter is TATA-less.

The predicted LTAg proteins from the Zambian bat PyVs contained a number of known functional motifs (Fig. 4), including the highly conserved DNAJ domain (HPDKGG) present in all the Zambian PyVs and reference PyV strains. Similarly, the LXXCE retinoblastoma (pRb1)-binding motif was present in each of the four newly reported species of the genera Alphapolyomavirus and Betapolyomavirus. The N-terminal conserved region 1 (CR1) motif (LXXXL) was present from residues 13 to 17 in all Zambian bat PyVs and predicted monopartite nuclear localization signals (NLSs) were evident in the LTAg sequences together with viral DNA origin-binding sites (A1, B1 and B2) and conserved Zn-finger (CX2,CX2–H2,H) and ATPase (GPX, GKT and GX,VNLE) motifs. The presence of putative miRNA and agnoprotein ORFs within the genomes of the newly
described Zambian PyVs was investigated by homology searches with sequences from other PyVs, but no significant regions of conservation were detected.

Phylogenetic analysis of monophyletic clusters of bat-NHP PyVs

Five monophyletic clusters (I–V) containing PyV species characterized from bats and NHPs were identified based on the LTAg protein (Fig. 2). A PyV whole-genome tree (Fig. S1, available in the online Supplementary Material) was consistent with the LTAg phylogenies with respect to the monophyletic clusters I, III, IV and V. Cluster II, however, was notably absent from the whole-genome phylogeny and in cluster I the *Rhinolophus hildebrandtii* polyomavirus 1 described in the present study grouped with NHP PyVs in the whole-genome phylogeny whereas it clustered with bat PyV species in the LTAg phylogeny. Comparison of the LTAg phylogeny with those derived from STAg, and particularly for VP1 and VP2 (Figs S2–S4), revealed incongruent tree topologies that are largely irreconcilable with the LTAg region and whole-genome phylogenies. These findings, the different clustering of *Rhinolophus hildebrandtii polyomavirus 1* in LTAg and whole-genome phylogenies and the discrepancies between LTAg and other sub-genomic regions, prompted us to perform a recombination analysis of the novel and previously described PyV genomes to investigate the distribution of recombination events within the family Polyomaviridae.

Recombination analysis of PyV genomes

By using the Recombination Detection Program (RDP) [19] suite of recombination detection algorithms and available genomes from species of PyV and the newly described Zambian species of PyV (*n* = 79), we detected 48 signals of recombination events affecting the relationships of species of PyV (Table S1). The majority of recombination events affected the structural region: 25 out of 48 (52 %); with 19 out of 48 (40 %) present in the nonstructural region and 4 out of 48 in the NCCR (8 %). Of these 48 PyV recombination events, comprising 80 recombinant sequences, the vast majority were

### Table 1. Pairwise nucleotide sequence identities of Zambian bat PyV LTAg coding regions with bat, NHP and human PyVs

<table>
<thead>
<tr>
<th>Miniopterus schreibersii PyV 1</th>
<th>Miniopterus schreibersii PyV 2</th>
<th>Rhinolophus hildebrandtii PyV 1</th>
<th>Rousettus aegyptiacus PyV 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Miniopterus schreibersii</em> PyV 1 (LC185215)</td>
<td>-</td>
<td>76</td>
<td>44</td>
</tr>
<tr>
<td><em>Miniopterus schreibersii</em> PyV 2 (LC185216)</td>
<td>76</td>
<td>-</td>
<td>43</td>
</tr>
<tr>
<td><em>Rhinolophus hildebrandtii</em> PyV 1 (LC185217)</td>
<td>44</td>
<td>43</td>
<td>-</td>
</tr>
<tr>
<td><em>Rousettus aegyptiacus</em> PyV 1 (LC185218)</td>
<td>41</td>
<td>42</td>
<td>61</td>
</tr>
<tr>
<td><em>Molossus molossus</em> PyV 1 (JKQ58893)</td>
<td>67*</td>
<td>67*</td>
<td>44</td>
</tr>
<tr>
<td><em>Artibeus planirostris</em> PyV 2 (JKQ58886)</td>
<td>66</td>
<td>66</td>
<td>45</td>
</tr>
<tr>
<td><em>Sturnira nicholsi</em> PyV 1 (JKQ58888)</td>
<td>66</td>
<td>65</td>
<td>44</td>
</tr>
<tr>
<td><em>Pan troglodytes</em> PyV 1 (FR692334)</td>
<td>58</td>
<td>58</td>
<td>47</td>
</tr>
<tr>
<td><em>Cebus albifrons</em> PyV 1 (JX159988)</td>
<td>44</td>
<td>45</td>
<td>65*</td>
</tr>
<tr>
<td><em>Dobsonia moluccensis</em> PyV 2 (AB972947)</td>
<td>42</td>
<td>42</td>
<td>61</td>
</tr>
<tr>
<td><em>Hunan PyV 13</em> (KF954417)</td>
<td>65</td>
<td>66</td>
<td>45</td>
</tr>
<tr>
<td><em>Otomops martienseni</em> PyV 1 (JX520658)</td>
<td>65</td>
<td>65</td>
<td>44</td>
</tr>
<tr>
<td><em>Pliocolobus rafomitarus</em> PyV 1 (JX159984)</td>
<td>65</td>
<td>65</td>
<td>45</td>
</tr>
<tr>
<td><em>Chlorocebus pygerythrus</em> PyV 3 (AB767297)</td>
<td>46</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td><em>Chlorocebus pygerythrus</em> PyV 1 (AB767298)</td>
<td>64</td>
<td>64</td>
<td>44</td>
</tr>
<tr>
<td><em>Acerodon celebensis</em> PyV 2 (AB972941)</td>
<td>44</td>
<td>43</td>
<td>64</td>
</tr>
<tr>
<td><em>Dobsonia moluccensis</em> PyV 3 (AB972946)</td>
<td>43</td>
<td>41</td>
<td>60</td>
</tr>
<tr>
<td><em>Saimiri sciureus</em> PyV 1 (JX159989)</td>
<td>43</td>
<td>43</td>
<td>65*</td>
</tr>
<tr>
<td><em>Mastomys natalensis</em> PyV 1 (AB858640)</td>
<td>42</td>
<td>43</td>
<td>63</td>
</tr>
<tr>
<td><em>Artibeus planirostris</em> PyV 2 (JQ958868)</td>
<td>44</td>
<td>43</td>
<td>65*</td>
</tr>
<tr>
<td><em>Dobsonia moluccensis</em> PyV 3 (JX520658)</td>
<td>42</td>
<td>43</td>
<td>60</td>
</tr>
</tbody>
</table>

*The percentage similarity to the most closely related PyV species is shown in bold type.*
detected in PyVs from NHPs, bats and humans (70/80; 87.5\%) (Table S1). NHP PyVs comprised 37.5\% (30/80) of recombinant sequences; with 31.25\% (25/80) detected in bat PyVs, including all newly reported bat PyV species in the present study. Human PyV species represented 18.75\% (18/80) of recombinant PyV sequences.

Fig. 1. Schematic diagram showing the genomic organization of Zambian insectivorous and fruit bat PyV species. Predicted coding regions for STAg (small T antigen), ALTO (alternate frame of the large T open reading frame), LTAg (large T antigen) and the variable protein (VP) 1–3 genes are indicated.
Table 2. Genome features of PyVs from Zambian insectivorous and fruit bat species

<table>
<thead>
<tr>
<th>Virus</th>
<th>Accession number</th>
<th>Genus</th>
<th>Cluster</th>
<th>Genome size (nt)</th>
<th>DNA G+C content (%)</th>
<th>ORF (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miniopterus schreibersii</td>
<td>LC185213</td>
<td>Alphapolyomavirus</td>
<td>V</td>
<td>4752</td>
<td>55.9</td>
<td>STAg</td>
</tr>
<tr>
<td>Miniopterus schreibersii</td>
<td>LC185214</td>
<td>Alphapolyomavirus</td>
<td>V</td>
<td>4878</td>
<td>55.8</td>
<td>LTAg</td>
</tr>
<tr>
<td>Miniopterus schreibersii</td>
<td>LC185215</td>
<td>Alphapolyomavirus</td>
<td>V</td>
<td>4877</td>
<td>55.7</td>
<td>ALTO</td>
</tr>
<tr>
<td>Rousettus aegyptiacus</td>
<td>LC185216</td>
<td>Betapolyomavirus</td>
<td>I</td>
<td>5060</td>
<td>38.5</td>
<td>VP1</td>
</tr>
<tr>
<td>Rhinolophus hildebrandtii</td>
<td>LC185217</td>
<td>Betapolyomavirus</td>
<td>I</td>
<td>5014</td>
<td>54.2</td>
<td>VP2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VP3</td>
</tr>
</tbody>
</table>

**Fig. 2.** Bayesian phylogenetic analysis of the LTag protein of PyVs. Virus names and GenBank accession numbers are shown. Zambian bat PyV LTag sequences are coloured in red and LTag sequences from other taxa previously identified from bats, non-human primates and humans are highlighted in pink, green and blue, respectively. Red circles at tree nodes indicate PyV genera and the monophyletic clusters are indicated at the right by Roman numerals. Bar, 0.4 amino acid substitutions per site. Centropristis striata polyomavirus 1 (KP071318) was employed as an outgroup.
Recombination signals were detected in five (6.25%) other mammalian PyV species (dolphin, raccoon, sea lion, badger and mouse). Finally, the remainder consisted of signals from bird PyVs (4/80; 5%) and a single recombinant sequence from a fish PyV (1/80; 1.25%). Despite the high statistical support for the detected recombination events, the recombination breakpoints are difficult to define, as the identities of the major and minor parental sequences were unknown in the majority of cases (Table S1) which is likely to be due to the ancient nature of these recombination events and subsequent mutations leading to low level of identity shared between sequences of extant species of PyV [2].

As Bayesian methods are more sensitive for detecting the effects of recombination events [20], in order to test the topological effects of the putative recombination events in the taxa containing the novel bat PyV sequences, we calculated the marginal likelihood of the five monophyletic clusters (Fig. 2) as topological hypotheses along the aligned genomes (Fig. S5). Strikingly, the marginal likelihood of the Bayesian tests supported the monophyletic clusters along the complete genomes for all clusters except cluster II, and to a lesser degree cluster III, within the VP1 region, which were polyphyletic in this structural gene region (Fig. S5).

We calculated a phylogenetic compatibility matrix, which compares how similar are the phylogenies inferred from two different sections of the genome [21], over all known species of PyV, including the newly described Zambian PyV genomes. Phylegetic incompatibilities between different regions of the genome are shown in Fig. 5. Two highly incompatible regions in the VP1 and NCCR are shown in red, indicative of possible recombination hot spots in the structural and promoter regions, respectively. Two possible cold spots for recombination, depicted in yellow, are located within the LTAg region and also within the VP2 gene. Fig. 5 depicts the distribution of recombination boundaries and two potential recombination breakpoints within the PyV alignments with boundaries detected by RDP analysis, aligned with the map in the phylogenetic compatibility matrix panel in Fig. 5 above. Evidence for recombination events in the capsid structural gene region were highly supported by the recombination boundaries flanking the coding region. Specifically, the VP3/VP1 region was bounded by two significant (P<0.05) recombination boundaries at the
Fig. 4. Amino acid sequence alignment of the LTg N-terminus from the Zambian bat PyVs and related reference strains. The HPDKGGD (B1 domain), the LXCE (pRb1-binding domain), the TPKKK (predicted nuclear localization signal, NLS) and DNA origin-binding sites, SNX1 (A1), FISHR (B1), RHHVSA (B2), zinc finger motif (Zn-finger) and ATPase motifs are indicated. Prediction of importin α-dependent nuclear localization signals was performed with NLS Mapper: http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi.
Fig. 5. Recombination analysis of species of PyV. The horizontal axis shows the positions adjusted to *Rousettus aegyptiacus polyomavirus 1* as a reference and colour-coding of each panel is shown in the respective legends to the right. (a) Phylogenetic compatibility matrix over the species of PyV. Each cell in the matrix reflects the normalized Robinson–Foulds distance between two neighbour-joining (NJ) trees corresponding to the respective regions in the PyV alignment. The vertical axis corresponds to the positions adjusted to the genome of *Rousettus aegyptiacus polyomavirus 1*. (b) Distribution of recombination breakpoints within the PyV alignment. Boundaries detected by RDP analysis are noted at the top of the panel. The vertical axis is the number of recombination boundaries per window. The solid black line shows the number of detected breakpoints. The grey and red areas indicate local 99 and 95% confidence intervals for the breakpoint clustering, respectively.
3’ end of the VP2/3 and also within the intergene region between structural (VP1) and nonstructural (LTAg) regions encompassing the incompatible region identified in Fig. 5. In contrast, the early region represented a cold spot for recombination events with evidence of low to no gene flow within the LTAg gene.

**DISCUSSION**

Consistent with results from previous evolutionary analyses of American, African and Indonesian bat PyVs, and despite the limited number of species analysed, the whole-genome nucleotide and LTAg protein phylogenies of Zambian bat PyVs were also paraphyletic (dispersed across the phylogenies) which is not inconsistent with a hypothesis that horizontal gene flow has occurred between viruses infecting mammalian species. A Bayesian-inferred phylogeny of the LTAg with all available PyV genomes and the newly described species of PyV from Zambian bats showed congruence with the genome analysis, with five monophyletic clusters of species of PyVs from bats and NHP species evident, with the exception of the monophyletic cluster II which was absent in the genome analysis. In addition, cluster I, containing the newly described *Rhinolophus hildebrandtii polyomavirus 1*, segregated with NHP PyVs in the whole-genome tree and with bat PyVs in the LTAg phylogeny. The phylogenies of the STAg, VP1 and VP2 regions were incompatible with those of the whole genomes and LTAg, indicative of gene flow/recombination events. This finding of congruence between the topologies of LTAg and whole-genome trees and the evidence for a cold spot for recombination in the LTAg identified in the present study strongly support the proposal for the use of this nonstructural gene region to determine evolutionary relationships and to identify novel species of PyVs based on genetic distances in this region [3]. This clustering of bat and NHP PyVs is also in agreement with the hypothesis proposed by Tao and co-workers [5] that host-switching may have occurred over evolutionary time. However, the recombination analysis in this previous study and also in the present work show that, due to the probable ancient nature of the gene flow, the parental major and minor species and direction are difficult to ascertain with all currently available data. This is potentially attributable to genetic drift over evolutionary time that decreases the shared nucleotide sequence identity between species of extant species of PyV [20].

Due to the absence of single-taxa clustering of bat, NHP and human PyVs, it has been previously suggested that bat PyVs are an unlikely source of human infection [5, 11]. The recent discovery of *Human polyomavirus 13* (New Jersey polyomavirus) in an immunocompromised pancreatic transplant recipient who was affected by the aftermath of, and breakdown in sanitation after, the Hurricane Sandy disaster [17] indicates that the recognized PyV whole genomes from sampled species may not entirely reflect all the gene flow events. Specifically, the segregation of *Human polyomavirus 13* with NHP and bat PyVs (including the two newly described *Miniopterus* PyVs identified in the present study) indicates that further sampling and genetic characterization of PyVs from human, NHP, bat and, particularly, other mammalian reservoirs (including rodents, see below) is required to help delineate the full extent of the evolutionary relationships and recombination events. It is noteworthy in this regard, and suggestive of ongoing horizontal gene flow between mammalian hosts, that *Miniopterus schreibersii polyomavirus 1* and 2 within monophyletic cluster V shared 64% identity across the LTAg coding region with *Chlorocebus pygerythrus polyomavirus 1* [15] identified in Mfuwe, Zambia from a vervet monkey and similarly, in cluster I, *Rhinolophus hildebrandtii polyomavirus 1* and *Rousettus aegyptiacus polyomavirus 1* exhibited 63 and 62% identity, respectively, across the LTAg gene with *Mastomys natalensis polyomavirus 1* from a multimammate mouse identified in Namwala in Zambia [18].

Fruit bats are known to harbour a number of high-consequence pathogens and the Egyptian fruit bat *Rousettus aegyptiacus* is considered a likely natural reservoir host for *Marburgvirus* [22] as entry into caves and mines populated by this bat has been associated with outbreaks of filoviral hemorrhagic fever [23, 24]. The detailed mechanism for precisely how *Marburgvirus* first transmits from its animal host to humans is unclear and numerous factors are probably involved in the sporadic outbreaks of severe haemorrhagic disease in sub-Saharan Africa. Interestingly, a real-time PCR analysis has indicated the potential for oral, urine, fecal or sexual transmission in distinct pulses of virus infection in older juvenile bats that coincides with the peak of the biannual birthing seasons [25]. Whether filoviral transmission between *Rousettus aegyptiacus* colonies is positively or negatively affected upon by co-infection and replication of other agents, including PyVs, is unknown and whether this could influence spillover into human populations deserves further study.

In *Human polyomavirus 5* (Merkel cell polyomavirus, MCPyV), ALTO in the early region, has been identified as an alternative transcript of the LTAg ORF [26]. ALTO is expressed during, but not necessary for, viral genome replication of the MCPyV genome. Previously, ALTO has been described in PyVs in bats (*Tadarida brasiliensis polyomavirus 1* and 2) and *Otomops martiensseni polyomavirus 2* [11]. A putative ALTO gene was identified in the Alphapolyomaviruses, *Miniopterus schreibersii polyomavirus 1* and 2 and shorter versions in the novel Zambian Betapolyomaviruses.

Genetic insertion and deletion events within the highly variable NCCR have been implicated as a primary mechanism for adaptation to the host [27]. Cell specificity of PyVs is thought to be the result of restricted tissue distribution of transcription factors required for viral gene expression. We identified a 126 bp deletion within the NCCR of one of the three strains of *Miniopterus schreibersii polyomavirus 1* and each PyV was characterized from splenic tissue. Further work is required to determine the effects on cellular
localization and gene expression driven by the wild-type and this deletion variant of the NCCR.

Despite the relatively small numbers analysed in the present study, high PyV prevalence (over 30%) was detected in bat species of the genera *Miniopterus*, *Rhinolophus* and *Rousettus* from Zambia. This high prevalence of PyV in bat species has also been seen in a previous survey of PyV diversity in Central American (Guatemalan) and African (Kenyan) bat species [5] employing the same semi-nested pan-PyV PCR used in the present study, where *Otomops martiensseni*, species of the genus *Chaerophon* (family *Molossidae*) and *Eidolon helvum* (family *Pteropodidae*) were found to have the highest prevalence rates (>20%). The elevated prevalence of PyVs is also seen in Indonesian fruit bats, with over 8% of species of the genera *Acrodon*, *Pteropus* and *Dobsonia* being positive for PyV DNA [10].

Tao and colleagues previously described three well-supported monophyletic clusters of PyVs from bats and NHPs employing partial sub-genomic early and late gene regions and suggested that this arose from multiple PyV jumps between bats and NHPs. One caveat arising from previous work and the present study in assigning clades with monophyletic bat/NHP PyV species and recombination analyses is of sampling bias as the majority (58%) of the formally recognized species of PyV are from either NHPs (n=23) or bats (n=21). Excluding PyVs from NHPs, bats and humans (n=13), only 11 other mammalian PyV genomes have been annotated with only four rodent (three mouse and one hamster) PyVs described. The remainder comprise a single fish PyV and seven bird PyV genomes. A seminal study by Buck and co-workers have recently described diverse species of PyV from fish, insects and arachnids and provided a theoretical framework for understanding of the evolution of the family *Polyomaviridae* [2].

The authors suggest the best virus–host co-evolutionary model to explain the observed diversity of PyVs is one of intrahost divergence where viruses diverge within a host faster than host animal speciation events. The identification of closely related bat PyVs within a single host [5, 8–11] is explained within this model by ancient viral divergence events within a single host species giving rise to distinct but closely related viral species, as seen in herpesviruses for example, and *Miniopterus schreibersii polyomavirus 1* and 2 in the present study.

Partial (approximately 300 nt) PyV VP1 fragments from *Rousettus aegyptiacus* have been previously described which segregated into two clades, *Rousettus PyV 1* and *Rousettus PyV 2* [5]. The present study reports the first PyV whole genome, to our knowledge, characterized from the Egyptian fruit bat *Rousettus aegyptiacus* and provides annotation of the genomes of an additional three novel PyVs in Zambian insectivoros bat species of the genera *Miniopterus* and *Rhinolophus*. Four of the six genomes were less than 5 kb in length: *Miniopterus schreibersii polyomavirus 1a* (4752 bp), *Miniopterus schreibersii polyomavirus 1b* and 1c (4878 bp) and *Miniopterus schreibersii polyomavirus 2* (4877 bp). The 21 previously described bat PyVs that have been designated as species [3] have genomes ranging in size from 4903 to 5372 bp (mean genome length: 5123 bp). Bat PyV members of the genera *Alphapolyomavirus* (n=12) and *Betapolyomavirus* (n=9) have similar ranges and mean genome lengths (in parentheses): 4903–5372 bp (5135 bp) and 5019–5213 (5107 bp), respectively. The previously described *Miniopterus africanus polyomavirus 1* genome is also 5213 bp in length; however, this virus is a Betapolyomavirus whereas the newly described Zambian species of PyV from *Miniopterus schreibersii* are Alphapolyomaviruses. The *Tadarida brasiliensis polyomavirus 1* and 2 genomes are 4882 and 4893 bp, respectively [11]; however, they are not yet classified as species by the *Polyomaviridae* Study Group of the International Committee on Taxonomy of Viruses. Two other recently described PyV genomes that exemplify the range of known diversity in PyV genome sizes are noteworthy: the PyV from giant guitarfish (*Rhynchobatus djiddensis*) (3962 bp) and *Centropristis striata polyomavirus 1* (sea bass PyV; 7369 bp) [2]. Taken together, these findings indicate that the archetypally sized approximately 5.1–5.2 kb *Human polyomavirus 1* (BKPyV) and *Macaca mulatta polyomavirus 1* (SV40 PyV) genomes may represent a median in the range of diversity of genome lengths and that greater variation exists; however, increased sampling of the diversity of PyV genomes in animal lineages is required to better understand horizontal gene flow in polyomaviral evolution.

There was a non-random distribution of recombination breakpoints with hot spots in structural regions, notably in the intergene region between VP1 and LTAg and also at the 3’ end of VP2/3 with cold spots for recombination within early nonstructural gene regions in the LTAg coding region. We consider that the parental major and minor species cannot be readily delineated as the recombination events are so distant evolutionarily that the signal is not as strong as, for example, for adenoviruses or other stable dsDNA viruses [28]. We speculate that intergene recombination of structural regions between PyVs encompassing the capsid proteins, such as VP1, may have facilitated altered tropism and species-jumping events. A similar phenomenon of non-random recombination events with hot spots enriched in structural regions and stable nonstructural cold spots has been seen in RNA viruses, for example within the capsid-coding region, P1, of the picornaviruses [21]. The cold spot for recombination events detected within the LTAg non-structural gene region indicates that breakpoints occurring inside this region have been selected against during polyomaviral evolution as they probably have a deleterious effect on genome replication.

**METHODS**

**Sample collection and genomic DNA extraction**

Bats were collected from the Leopard’s Hill Cave in Lusaka, Zambia (15° 36.132’ S 28° 43.457’ E) in 2012. Ethical approval was provided by the then Zambia Wildlife
Author (ZAWA), now the Department of National Parks and Wildlife, Ministry of Tourism and Arts, to undertake the present study. Genomic DNA was extracted from bat spleen tissue using the QIAamp DNA minikit (Qiagen) according to the manufacturer’s instructions.

**PyV broad-spectrum and inverse PCRs**

The partial PyV VP1 and VP3–VP1 fragments were amplified employing a nested PCR strategy with degenerate broad-spectrum oligonucleotides as previously described [14]. The primary PCR was performed with 100 ng of bat spleen genomic DNA and a proof-reading Taq DNA polymerase, High Fidelity PCR Master kit (Roche) in 20 μl final volumes with the following thermocycling conditions: 1 cycle of 94 °C for 2 min followed by 45 cycles of 94 °C for 30 s, 56 °C for 1 min and 72 °C for 1 min. The secondary PCR was performed by addition of 0.5 μl of the primary PCR product under the following thermocycling conditions: 1 cycle of 95 °C for 5 min followed by 45 cycles of 94 °C for 30 s, 56 °C for 30 s and 72 °C for 1 min with a final extension of 72 °C for 1 min. Amplicons of approximately 250 bp were visualized by electrophoresis on 1.5 % (w/v) agarose gels and direct sequencing of the PCR amplicons was performed on an ABI Prism 3100 Genetic Analyzer using Big Dye v3.1 chemistry (Applied Biosystems). Full-length genomes were amplified from bat genomic DNA by inverse PCR employing the Tsks GFlex DNA polymerase (Takara) using contiguously oppositely facing oligonucleotide primers with the following thermocycling conditions: 1 cycle of 94 °C for 2 min followed by 45 cycles of 94 °C for 30 s, 46 °C for 1 min and 72 °C for 1 min. The secondary PCR was performed by addition of 0.5 μl of the primary PCR product under the following thermocycling conditions: 1 cycle of 95 °C for 5 min followed by 45 cycles of 94 °C for 30 s, 56 °C for 30 s and 72 °C for 1 min with a final extension of 72 °C for 1 min. Amplicons of approximately 5 kb were cloned into the pCR-Blunt II-TOPO vector (Invitrogen) and genome assembly was performed by primer walking. Species designations were obtained by amplification and sequencing of the bat mitochondrial *cytochrome b* gene as previously described [13].

**Phylogenetic and recombination analyses**

The complete genomes of PyVs, nucleotide and amino acid sequences of proteins were multiple sequence aligned with MAFFT using the algorithm FFT-NS-1 [29]. Phylogenetic trees were inferred using MrBayes v3.2 performing chains with sufficient generations to ensure the standard deviations of split frequencies were less than 0.01 [30]. The general time-reversible substitution model with gamma-distributed rates across sites (GTR+G+I) was assumed for the nucleotide alignment and the sliding window analysis used in the topology hypothesis testing of the clusters. Analyses over proteins used a mixed substitution model to explore the best substitution model for each protein. The multiple genome alignment was analysed with the suite of recombination detection algorithms in RDP v4.58 [19] for evidence of recombination events between species of PyV. The considered methods were RDP [31], GENECONV [32], Chimaera [33], MaxChi [34], Bootscan [35], SiScan [36] and 3Seq [37], with a P threshold <0.01 and requiring the signal to be detectable simultaneously by at least three methods. To confirm the likelihood of recombination events distorting the phylogenetic relationship of the identified monophyletic clusters along regions of the genome, a sliding window analysis with 500 nucleotides and 250 nucleotides per step was performed with likelihood estimation with a stepping-stone approach in MrBayes [38]. The effects of recombination events in the phylogeny of sections of the genome was assessed with a compatibility matrix similar to that described previously by Heath and colleagues [21], where the compatibility of two windows with 500 nucleotides from a sliding window and 50 nucleotides per step is defined as the normalized Robinson–Foulds distance [39] between the corresponding neighbour-joining phylogenetic trees under a Tamura–Nei evolutionary model; thus, the compatibility reflects how similar are the phylogenies inferred for any two genome windows ranging from 0 (identical) to 1 (completely dissimilar). We also included the analysis of the recombination breakpoints, using RDP 4 with a sliding window of 500 nucleotides with a one nucleotide step and 1000 permutations for estimating the statistical support of the breakpoint distribution.

**Funding information**

This study was supported in part by grants to Professor Hirofumi Sawa for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Science, Sports and Technology (MEXT) of Japan (16H06429, 16H06431, 16K21723), grants from the Ministry of Education, Culture, Sports, Science and Technology, the Ministry of Health, Labour and Welfare, Japan (MEXT)/JSPS KAKENHI (16H05805); and the Japan Initiative for Global Research Network on Infectious Diseases (J-GRID), from the Japan Agency for Medical Research and Development (AMED).

**Acknowledgements**

The authors would like to thank the Department of National Parks and Wildlife of the Ministry of Tourism and Arts (formerly ZAWA) for assistance with sample collection in Zambia.

**Conflicts of interest**

The authors declare that there are no conflicts of interest.

**Ethical statement**

Ethical approval was provided by the then Zambia Wildlife Authority (ZAWA), now the Department of National Parks and Wildlife, Ministry of Tourism and Arts, to undertake the present study.

**References**

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.