

# Discovery of diverse polyomaviruses in bats and the evolutionary history of the *Polyomaviridae*

Ying Tao,<sup>1†</sup> Mang Shi,<sup>2†</sup> Christina Conrardy,<sup>1</sup> Ivan V. Kuzmin,<sup>3</sup> Sergio Recuenco,<sup>3</sup> Bernard Agwanda,<sup>4</sup> Danilo A. Alvarez,<sup>5</sup> James A. Ellison,<sup>3</sup> Amy T. Gilbert,<sup>3</sup> David Moran,<sup>5</sup> Michael Niezgodna,<sup>3</sup> Kim A. Lindblade,<sup>6</sup> Edward C. Holmes,<sup>2,7</sup> Robert F. Breiman,<sup>8</sup> Charles E. Rupprecht<sup>3</sup> and Suxiang Tong<sup>1</sup>

## Correspondence

Suxiang Tong  
sot1@cdc.gov

<sup>1</sup>Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

<sup>2</sup>Sydney Emerging Infections and Biosecurity Institute, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia

<sup>3</sup>Division of High Consequence Pathogens and Pathology, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA

<sup>4</sup>National Museum, Kenya Wildlife Service, Nairobi, Kenya

<sup>5</sup>Center for Health Studies, Universidad del Valle de Guatemala, Guatemala City, Guatemala

<sup>6</sup>Centers for Disease Control and Prevention Central America and Panama, Guatemala

<sup>7</sup>Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA

<sup>8</sup>Centers for Disease Control and Prevention Kenya, Nairobi, Kenya

Polyomaviruses (PyVs) have been identified in a wide range of avian and mammalian species. However, little is known about their occurrence, genetic diversity and evolutionary history in bats, even though bats are important reservoirs for many emerging viral pathogens. This study screened 380 specimens from 35 bat species from Kenya and Guatemala for the presence of PyVs by semi-nested pan-PyV PCR assays. PyV DNA was detected in 24 of the 380 bat specimens. Phylogenetic analysis revealed that the bat PyV sequences formed 12 distinct lineages. Full-genome sequences were obtained for seven representative lineages and possessed similar genomic features to known PyVs. Strikingly, this evolutionary analysis revealed that the bat PyVs were paraphyletic, suggestive of multiple species jumps between bats and other mammalian species, such that the theory of virus–host co-divergence for mammalian PyVs as a whole could be rejected. In addition, evidence was found for strong heterogeneity in evolutionary rate and potential recombination in a number of PyV complete genomes, which complicates both phylogenetic analysis and virus classification. In summary, this study revealed that bats are important reservoirs of PyVs and that these viruses have a complex evolutionary history.

Received 12 September 2012

Accepted 09 December 2012

## INTRODUCTION

Polyomaviruses (PyVs) are small DNA viruses of the family *Polyomaviridae*. Due to improvements in molecular diagnostic techniques, a number of PyVs have been recently discovered. Ten human PyVs have been identified to date: BK virus (BKPyV), JC virus (JCPyV), KI PyV

(KIPyV), WU PyV (WUPyV), Merkel cell PyV (MCPyV), human PyV 6 and 7 (HPyV6 and -7), *Trichodysplasia spinulosa*-associated PyV (TSPyV), human PyV 9 (HPyV9) (Van Ghelue *et al.*, 2012), and MW PyV (MWPyV) (Siebrasse *et al.*, 2012). Primary PyV infection in humans usually occurs in childhood and seemingly results in lifelong persistence. In healthy humans, PyVs have not been associated with acute disease. However, PyV reactivation in the case of BKPyV, JCPyV and MCPyV can cause severe disease in people with immunodeficiency (Jiang *et al.*, 2009). In addition to humans, PyVs have been found in a wide range of mammalian and avian hosts including primates, rodents, sea lions, cows, horses, rabbits, bats and

†These authors contributed equally to this work/paper.

The GenBank/EMBL/DDBJ accession numbers for the genome sequence of the bat polyomaviruses determined in this study are JX520641–JX520664.

Eight supplementary figures and four tables are available with the online version of this paper.

birds (Johne *et al.*, 2011; Norkin *et al.*, 2011; Renshaw *et al.*, 2012). The presence and variety of PyVs in bats is of particular interest, as this group of mammals is known to harbour and transmit a variety of emerging viruses (Kuzmin *et al.*, 2011). Until recently (Fagrouch *et al.*, 2012), only a single bat PyV species had been identified in bats (Misra *et al.*, 2009). To understand better the role of bats in the maintenance and transmission of PyVs, we undertook a more extensive surveillance and analysis.

The ORFs of the dsDNA genome of PyVs are separated by a non-coding control region (NCCR) into an early region that contains genes for regulatory proteins [large T-antigen (LT-Ag) and small T-antigen (ST-Ag)] and a late region that contains genes for structural proteins (VP1, VP2 and VP3) (Johne *et al.*, 2011). The bidirectional NCCR controls the transcription of both early and late promoters and regulates the initiation of viral DNA synthesis. The NCCR has the highest level of variation among PyVs due to rearrangement of the enhancer elements, including mutations, deletions and duplications, which may enable the host adaptation of certain PyVs (White *et al.*, 2009).

One of the key aspects of PyV evolution is that phylogenetic trees based on the early and late regions of PyV genomes show significant incongruence (Krumbholz *et al.*, 2009), especially for the recently discovered human viruses WUPyV, KIPyV, HPyV6 and HPyV7 (Allander *et al.*, 2007; Gaynor *et al.*, 2007; Schowalter *et al.*, 2010). Specifically, whilst the late regions of these human PyVs are extremely divergent from other PyVs, the early regions exhibit much less divergence. Similarly, discrepancies exist in the level of genetic divergence among genes (i.e. rate heterogeneity) in many PyV species (Allander *et al.*, 2007; Krumbholz *et al.*, 2009), although this has yet to be characterized systematically.

Whilst its exact cause is unclear, the incongruence between the phylogenies of the PyV early and late regions has complicated PyV classification. Under the initial classification scheme, the family *Polyomaviridae* contained a single genus. However, the International Committee on Taxonomy of Viruses (<http://ictvonline.org>) has recently proposed three genera: *Orthopolyomavirus*, *Wukipolyomavirus* and *Avipolyomavirus* (Johne *et al.*, 2011). The genus *Wukipolyomavirus* includes all those viruses (KIPyV, WUPyV, HPyV6 and HPyV7) with highly divergent VP1 and VP2 genes. However, this classification scheme is not compatible with the phylogenetic history of the viruses, as the KIPyV–WUPyV cluster does not group with HPyV6 and HPyV7 in the early-gene regions (Fig. S1, available at JGV Online).

It is widely accepted that DNA viruses that establish persistent infections sometimes co-diverge with their hosts on timescales of many millions of years. The idea of co-divergence has also been proposed for PyV, although not without controversy. PyV co-divergence was proposed originally based on the topological resemblance between the virus and host phylogenies (Shadan & Villarreal, 1993; Soeda *et al.*, 1980), with one study providing statistically

significant evidence for this specific virus–host relationship (Pérez-Losada *et al.*, 2006). However, after newly identified PyVs were included in phylogenetic analyses, the topology of the virus trees did not bear a significant resemblance to that of the host, and co-divergence was rejected using statistical tests (Krumbholz *et al.*, 2009).

In this study, we examined 380 specimens from 35 bat species from Kenya ( $n=195$ ) and Guatemala ( $n=185$ ) for the presence of PyVs, and characterized the PyV genomes with the particular aim of determining the genetic diversity of the family *Polyomviridae*, the pattern and cause of any phylogenetic incongruence among them and the extent of virus–host co-divergence.

## RESULTS

### Prevalence and diversity of PyVs in bats

The specimens consisted of 195 rectal swabs from 195 Kenyan bats representing 22 different bat species, and 91 rectal and 94 oral swabs from 96 Guatemalan bats representing 13 different bat species. Overall, PyV DNA was detected in 23 Kenyan bats (11.8%) and one Guatemalan bat (1.0%). Of the 35 bat species examined in this study, seven harboured PyVs: *Otomops martiensseni* (six of 19), *Chaerephon* sp. (eight of 35), *Eidolon helvum* (two of nine), *Rousettus aegyptiacus* (five of 46), *Cardioderma cor* (one of 14), *Miniopterus africanus* (one of one) and *Pteronotus davyi* (one of 17) (Table 1). Due to a lack of available reference sequences, we were not able to determine the exact species of the *Chaerephon* PyVs. Among the bats surveyed, *O. martiensseni* and *Chaerephon* sp. from the family Molossidae, and *E. helvum* from the family Pteropodidae had a high PyV prevalence with at least 20% positive rates. PyV was detected from the one *M. africanus* bat surveyed, but there was an insufficient number of samples to infer prevalence.

Phylogenetic analyses were performed to determine the evolutionary relationships of these newly identified bat PyVs (Figs 1 and 2). Prior to phylogenetic inference, viral taxa shown to experience significant rate heterogeneity (see below) were excluded. Specifically, KIPyV, WUPyV, HPyV6 and HPyV7 were excluded from the VP1 and VP2 trees, and all avian PyVs were excluded from the LT-Ag tree. For comparative purposes, the phylogenies showing these highly divergent lineages are shown in Fig. S1.

Based on the partial VP1 phylogeny, the 24 bat PyVs identified in our study formed 12 distinct lineages (Fig. 1). Although these 12 lineages were phylogenetically distinct, they fell into four clusters: (i) *Cardioderma* PyV, *Otomops* PyV1, *Eidolon* PyV 1 and *Rousettus* PyV1 grouped with MCPyV and related PyV species; (ii) *Chaerephon* PyV1 and *Otomops* PyV2 grouped with ChPyV; (iii) *Chaerephon* PyV2, *Eidolon* PyV2 and *Otomops* PyV3 formed an orphan clade whose relationship with other PyV species remained uncertain; and (iv) the remaining lineages (*Pteronotus* PyV,

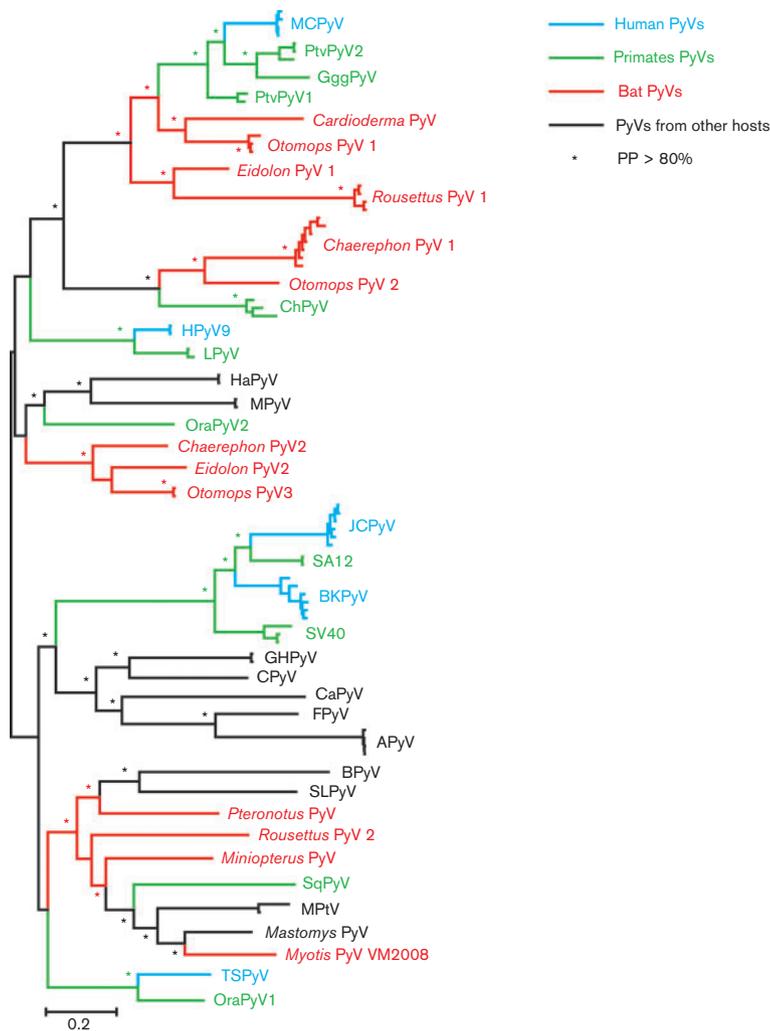
**Table 1.** Detection of PyV in bats from Kenya and GuatemalaSee Tong *et al.* (2009, 2012) for location information. A question mark indicated unknown location.

Bat species	Location [virus detection rate (no. positive/no. tested)]
<b>Kenya</b>	
<i>Cardioderma cor</i>	15 (0/10); 12 (0/3); 6 (1/1)
<i>Chaerephon pumilus</i>	11 (0/5)
<i>Chaerephon sp.</i>	6 (5/12); 17 (2/16); 3 (0/6); ? (1/1)
<i>Coleura sp.</i>	14 (0/1)
<i>Coleura afra</i>	11 (0/1)
<i>Eidolon helvum</i>	4 (2/9)
<i>Epomophorus wahlbergi</i>	9 (0/4)
<i>Hipposideros commersoni</i>	14 (0/9)
<i>Hipposideros ruber</i>	5 (0/2); 2 (0/4)
<i>Miniopterus africanus</i>	10 (1/1)
<i>Miniopterus inflatus</i>	5 (0/7)
<i>Miniopterus natalensis</i>	1 (0/1)
<i>Otomops martiensseni</i>	7 (6/19)
<i>Pipistrellus sp.</i>	14 (0/5)
<i>Rhinolophus hildebrandtii</i>	10 (0/4)
<i>Rhinolophus sp.</i>	13 (0/1); 14 (0/1); ? (0/2)
<i>Rhinolophus clivosus</i>	8 (0/3)
<i>Rousettus aegyptiacus</i>	13 (3/11); 5 (1/10); 1 (0/10); 2 (0/9); 16 (0/5); ? (1/1)
<i>Scotoecus sp.</i>	13 (0/16)
<i>Taphozous hildegardeae</i>	14 (0/2)
<i>Taphozous sp.</i>	11 (0/2); ? (0/1)
<b>Guatemala</b>	
<i>Artibeus jamaicensis</i>	El Naranjo (0/3); El vina (0/3); El Penate (0/1); El Jobo (0/4); Los Hilos (0/1)
<i>Artibeus lituratus</i>	Los Tarrales (0/1)
<i>Carollia castanea</i>	El Naranjo (0/1)
<i>Carollia perspicillata</i>	El Naranjo (0/1); Los Tarrales (0/1)
<i>Desmodus rotundus</i>	El Naranjo (0/11); Los Tarrales (0/1); El Penate (0/1); El Jobo (0/3); El Vina (0/6)
<i>Glossophaga soricina</i>	El Naranjo (0/3); Los Tarrales (0/1); El Penate (0/1); El Jobo (0/3)
<i>Micronycteris microtis</i>	El Naranjo (0/1); Los Hilos (0/5); El Jobo (0/2)
<i>Myotis nigricans</i>	El Jobo (0/1)
<i>Myotis sp.</i>	El Naranjo (0/1)
<i>Phyllostomus discolor</i>	Los Tarrales (0/2)
<i>Platyrrhinus helleri</i>	Los Tarrales (0/1)
<i>Pteronotus davyi</i>	El Naranjo (1/16); Los Tarrales (0/1)
<i>Sturnira lilium</i>	Los Tarrales (0/4); El vina (0/3); El Penate (0/1); El Jobo (0/11); Los Hilos (0/1)

*Rousettus* PyV2 and *Miniopterus* PyV) grouped within the clade that contained BPyV, SLPyV, SqPyV, MPtV, *Mastomys* PyV and *Myotis* PyV VM2008. Whilst each lineage was associated with a single bat host species, *O. martiensseni*, *Chaerephon* sp., *E. helvum* and *R. aegyptiacus* bats were infected with PyVs from two or more distinct lineages.

From the 12 lineages, we were able to obtain full-genome sequences for PyVs from seven bat samples: KY397, KY157, KY156, KY270, KY336, KY369 and GTM203, representing lineages of *Chaerephon* PyV1, *Otomops* PyV1, *Otomops* PyV2, *Eidolon* PyV1, *Cardioderma* PyV, *Miniopterus* PyV and *Pteronotus* PyV, respectively. Each of the seven genomes and their ORFs demonstrated <80% nucleotide identity to the known reference PyVs (Table 2). Such a high level of divergence suggests that these seven bat PyV

lineages probably represent novel PyV species (Johns *et al.*, 2011). Although full-genome sequences were not available for the remaining five bat PyV lineages, their divergent partial VP1 sequences suggested that they also probably represent novel PyV species (Fig. 1). We compared the LT-Ag phylogeny from the late region with the VP1+2 phylogeny from the early region and observed a similar topology for six bat PyVs (Fig. 2). The exception was GTM203, which grouped with the JCPyV–BKPyV–SA12–SV40 cluster in the LT-Ag tree but with SLPyV based on the VP1+2 tree. Notably, none of the bat PyVs identified in this study showed a close relationship to *Myotis* PyV VM2008 identified from *Myotis* bats in Canada (Misra *et al.*, 2009) (Figs 1 and 2). Importantly, the bat PyV lineages were clearly paraphyletic (i.e. fell at multiple locations across the tree), indicative of multiple cross-species transmission events.

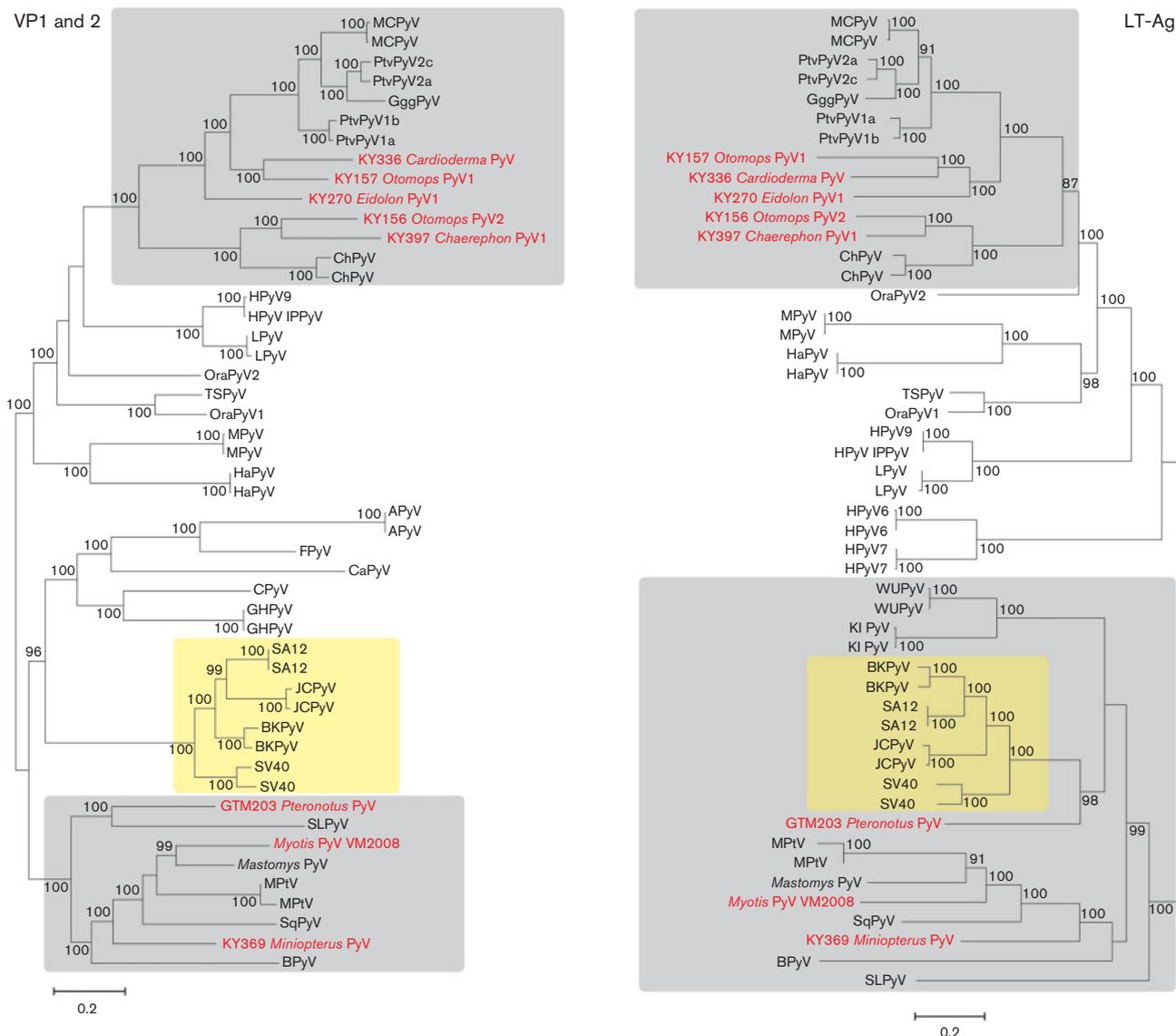


**Fig. 1.** Phylogenetic tree of partial VP1 genes of the bat viruses (represented by red branches) inferred using MrBayes v3.2. The names of lineages or PyV species are indicated. Posterior probability values (PP) >80% are indicated as asterisks above branches. Bar, number of nucleotide substitutions per site. APyV, Avian polyomavirus; BPyV, bovine PyV; CaPyV, canary polyomavirus; ChPyV, chimpanzee PyV; CPyV, crow polyomavirus; FPyV, finch polyomavirus; GggPyV, *Gorilla gorilla gorilla* polyomavirus; GHPyV, goose hemorrhagic polyomavirus; HaPyV, hamster PyV; LPyV, B-lymphotropic polyomavirus; MPyV, murine PyV; MPtV, murine pneumotropic polyomavirus; OraPyV, orangutan PyV; PtvPyV, *Pan troglodytes verus* polyomavirus; SA12, baboon polyomavirus 1; SLPyV, California sea lion PyV; SqPyV, squirrel monkey PyV; SV40, simian virus 40.

Whilst this paper was under review, an additional eight bat-associated PyVs were documented in animals sampled in French Guiana (Fagrouch *et al.*, 2012). Although there was little overlap in host species or geographical area between the two studies, the phylogenies of both VP1 + 2 and LT-Ag (Fig. S2) revealed that these novel South American bat PyVs formed well-supported monophyletic groups with the bat lineages described here: (i) group A of the South American bat PyVs clustered with Guatemalan PyV GTM203; (ii) group B was nested within KY157, KY336 and KY270; and (ii) group C formed a sister clade to KY156 and KY397. No South American PyV was found to be clustered with KY369 or *Myotis* PyV VM2008. However, although the bat-associated PyVs tended to group together, there was still a substantial genetic distance among them. Indeed, with one exception, all bat lineages exhibited <80% identity in comparison with each other (Table S1). The single exception was GTM203, with a whole-genome identity of 81% compared with French Guianan bat PyV2b\_R266, which was thus less than the species-defining threshold (81–84%; Johne *et al.*, 2011).

## Genome organization of bat PyVs

The seven generated PyV genomes were 4899–5372 bp, with an overall G+C content of 41–43 mol% (Table S2). All seven genomes showed the typical PyV genome organization (Fig. S3), comprising one strand coding for regulatory proteins (ST-Ag and LT-Ag) and the other for structural proteins (VP1, VP2 and VP3). Between the two coding regions was the NCCR. Additional ORFs were not found in these genomes with the exception of KY369, which contained a putative ORF of 204 bp upstream of VP2 (Table S2). This putative ORF has also been identified in JCPyV, BKPyV, SV40, SA12, BPyV, APyV, SqPyV, SLPyV, ChPyV and *Myotis* PyV (Misra *et al.*, 2009; Van Ghelue *et al.*, 2012), but the ORF in KY369 had no sequence homology to these ORFs. The ORF upstream of VP2 in JCPyV, BKPyV, SV40 and SA12 encodes an agnoprotein, which has a critical role in the regulation of viral gene expression and replication and in the modulation of host-cell functions including cell-cycle progression and DNA repair (Khalili *et al.*, 2005). Further studies are needed to characterize the function of the putative protein in bat PyVs.



**Fig. 2.** Phylogenetic trees of the early (LT-Ag) and late (VP1 + 2) genes of the bat virus genomes (represented by red labels) inferred using MrBayes v3.2. Posterior probability values are shown above the branches. In these phylogenies, well-supported clades that contained the bat PyVs are shaded grey and JCPyV–BKPyV–SA12–SV40 clusters are shaded yellow. Bars, number of nucleotide substitutions per site.

The NCCR region of the bat PyVs showed a low degree of sequence similarity among bat PyVs and when compared with other known PyVs (<66 % nucleotide identity). Several conserved elements were identified in the replication origin, including an A/T-rich domain and several LT-Ag-binding sites (GAGGC and its reverse complement GCCTC) (Fig. S4). The LT-Ag-binding elements varied in number among the different PyV species and were usually in different arrangements. The six bat PyVs obtained in this study contained four to six copies of these elements, whilst GTM203 had only two. In KY397 and KY156, one copy of the GAGGC element was overlapped by one copy of the GCCTC element in a palindromic octamer (GAGGCCTC).

Similar overlapped palindromic octamers were also observed in the NCCR of other PyV species such as ChPyV, MCPyV, GggPyV, PtvPyV1, PtvPyV2, OraPyV1 and TSPyV.

Like all other known PyVs, the ST-Ag and LT-Ag in the early region in seven bat PyVs shared around 80 aa at the N terminus. The ST-Ags in JCPyV, BKPyV and SV40 contained a cysteine-rich motif (CX<sub>5</sub>CX<sub>7-8</sub>CXCX<sub>2</sub>CX<sub>21-22</sub>CXCX<sub>2</sub>CX<sub>3</sub>WFG) at the C-terminal end of the proteins. This motif was perfectly conserved in KY270 and GTM203, but the initial cysteine residue of the motif was not present within the other five bat PyVs (X<sub>5</sub>CX<sub>7-8</sub>CXCX<sub>2</sub>CX<sub>21-22</sub>CXCX<sub>2</sub>CX<sub>3</sub>WFG). The LT-Ags of bat PyVs from this study varied from 671 to 836 aa (Table S2). The length difference

**Table 2.** Similarity between novel bat PyVs and their closest relatives

PyVs	Closest relatives	Nucleotide identity (%)	Amino acid identity (%)				
			VP1	VP2	VP3	LT-Ag	ST-Ag
KY397	KY156	72	75	69	64	79	64
	ChPyV	67	69	60	55	70	52
KY156	KY397	72	76	69	64	79	64
	ChPyV	67	43	66	63	70	52
KY157	KY336	70	71	76	73	64	75
	KY270	65	73	60	57	60	56
	MCPyV	65	69	58	53	55	52
KY270	KY336	67	69	59	56	61	55
	KY157	65	73	60	57	60	56
	PtvPyV1a	65	72	56	52	59	59
	MCPyV	65	70	56	52	60	59
KY336	KY157	70	71	76	73	64	75
	KY270	67	69	59	56	61	55
	PtvPyV1a	66	71	61	57	60	54
	MCPyV	66	70	57	51	60	51
KY369	<i>Mastomys</i> PyV	65	72	58	55	56	52
	SqPyV	65	65	57	55	61	53
	<i>Myotis</i> PyV	64	72	55	52	54	45
	GTM203	64	71	49	52	63	55
GTM203	KY369	64	71	49	52	63	55
	SV40	62	57	40	45	60	56
	<i>Myotis</i> PyV	61	69	45	48	52	42
	<i>Mastomys</i> PyV	61	69	48	51	54	45

mainly resided within two regions: one at the highly variable region between the J domain and nuclease localization signal sequence (NLS) and the other at the C terminus (Fig. S5). The LT-Ag of KY270, KY336, KY157, KY356 and KY397 had 70–180 aa insertions that were rich in serine, glutamine and threonine. Similar insertions were found in their close phylogenetic relatives: MCPyV, GggPyV, PtvPyV1, PtvPyV2, MPyV, OraPyV2 and HaPyV. The LT-Ag protein of KY369 and GTM203 lacked an insertion between the J domain and NLS, but had an additional 37–49 aa at the C-terminal end, as in JCPyV, BKPyV, SA12 and SV40. The additional C-terminal amino acids in JCPyV, BKPyV, SA12 and SV40 are known encoding a host-range domain (Cantalupo *et al.*, 2005). Despite the length variation, the LT-Ag of bat PyVs exhibited a high degree of conservation for functional domains, including the Cr1, J, Cr2, NLS, Zn finger and ATPase-p53 binding domains, as indicated in Fig. S5 (Pipas, 1992).

Of the three capsid proteins (VP1, VP2 and VP3), the major structural protein VP1 contains the antigenic determinants and plays an essential role in the entry of virus into host cells (Johne *et al.*, 2011). The VP2 and VP3 proteins may be necessary to ensure specific encapsidation of the replicated PyV genome (Johne *et al.*, 2011). Like other PyVs, the ORFs of VP1, VP2 and VP3 from bat PyVs had low sequence identity and different lengths compared with each other and with other PyVs (Tables 2 and S2), but protein sequence alignments showed conserved functional

elements including the putative loop structure in VP1 (Fig. S6), VP1-binding domain, NLS and DNA-binding domain, although there were variations and exceptions (Fig. S7).

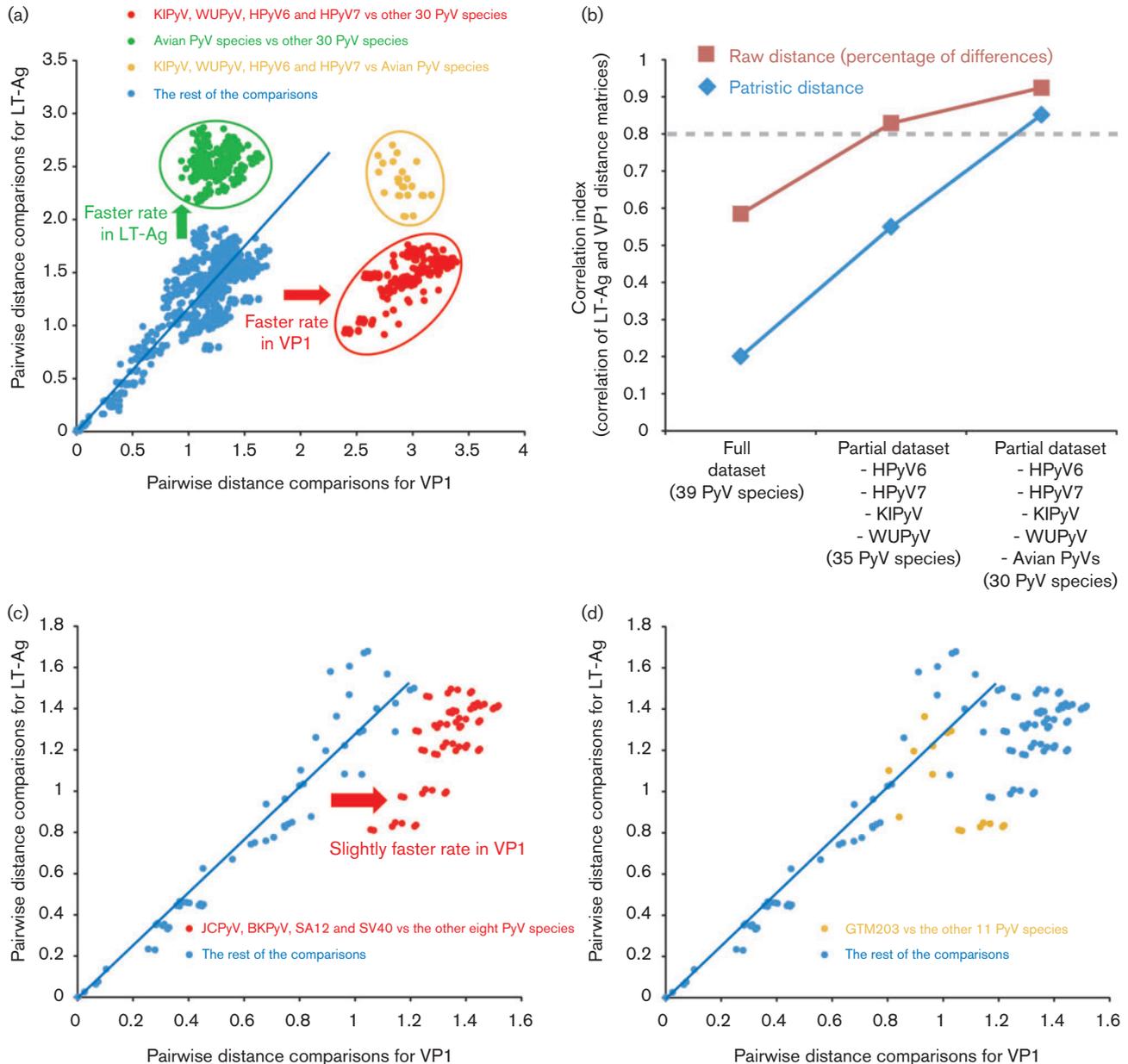
### Rate heterogeneity in PyVs

To determine the extent and pattern of rate heterogeneity in members of the family *Polyomaviridae*, we compared patristic genetic distance matrices among the VP1, VP2 and LT-Ag genes. Notably, significant rate heterogeneity was detected between the early and late gene matrices, and with extremely low correlation coefficients:  $r=0.200$  for VP1 compared with LT-Ag, and  $r=0.198$  for VP2 compared with LT-Ag. Conversely, no significant rate heterogeneity was detected between the two late genes (VP1 and VP2;  $r=0.848$ ). To identify which viral representatives were responsible for the low  $r$  value between early and late genes, we plotted the VP1 matrix against the LT-Ag matrix. Two groups of PyVs were identified as having rate heterogeneity (Fig. 3a): (i) KIPyV, WUPyV, HPyV6 and HPyV7 exhibited a substantial shift to the right, indicating that their VP1 genes have evolved significantly (three times or more) faster than those of other PyV species; and (ii) all avian PyV species exhibited an upward shift, indicating rate heterogeneity in the LT-Ag gene. To confirm rate heterogeneity of these groups, we excluded them from genetic distance comparisons and repeated the correlation analysis: the correlation between the VP1 and LT-Ag

matrices showed a significant improvement, thereby confirming the observation of rate variation (Fig. 3b).

In addition to rate heterogeneity, our phylogenetic analyses revealed a strong incongruence in evolutionary relationships between GTM203 and the JCPyV–BKPyV–SA12–SV40 cluster when trees based on different viral genes were compared (Fig. 2). In the LT-Ag phylogeny, GTM203

formed a monophyletic group with JCPyV–BKPyV–SA12–SV40, whereas in the VP1+2 phylogeny, they were distantly related. To determine whether this incongruence could in part be due to rate acceleration along specific lineages, we performed a rate heterogeneity analysis using seven closely related PyV species (SLPyV, *Myotis* PyV, *Mastomys* PyV, MPtV, SqPyV, *Miniopterus* PyV and



**Fig. 3.** Analyses of rate heterogeneity among members of the family *Polyomaviridae*. (a) Comparisons of genetic distance matrices (full dataset) between VP1 and LT-Ag. Each dot represents the comparison between one pair of PyVs: the x-axis value indicates their genetic distance in VP1, whilst the y-axis value is their distance in LT-Ag. Dots that exhibited a linear relationship are shown in blue, whilst those that deviated significantly from a linear relationship are indicated in other colours. (b) Comparisons of the correlation index,  $r$ , with or without the PyV species experiencing rate heterogeneity. The  $r$  values were calculated from both raw distance matrices and patristic distance matrices. (c, d) Comparisons of genetic distance matrices (partial dataset) to examine rate heterogeneity in the JCPyV–BKPyV–SA12–SV40 cluster (c) and in GTM203 (d).

BPyV). The comparisons involving GTM203 followed the general linear relationship (Fig. 3d), whereas those involving JCPyV, BKPyV, SA12 and SV40 fell slightly to the right (Fig. 3c). These results indicated that the phylogenetic incongruence was caused by the JCPyV–BKPyV–SA12–SV40 cluster, which may have evolved more rapidly in the late region of the genome and which in turn might lead to long-branch attraction (Bergsten, 2005). Indeed, in the VP1 + 2 tree, the JCPyV–BKPyV–SA12–SV40 cluster was separated from other known PyVs by a long branch (Fig. 2).

### Analysis of recombination in PyV

KIPyV, WUPyV, HPyV6 and HPyV7 were identified to be potential recombinants using six of the seven methods implemented in RDP. In contrast, JCPyV, BKPyV, SA12, SV40 and all of the avian PyVs were only identified by CHIMAERA (full results available from the authors on request). Hence, a signal for recombination was observed in all 13 PyV species shown to experience strong rate heterogeneity. Although these results are clearly compatible with the occurrence of recombination, it is important to note that, in every case, we could only identify one of the parental sequences. Indeed, phylogenetic tree inference based on the non-recombinant regions suggested that JCPyV, BKPyV, SA12, SV40, KIPyV, WUPyV, HPyV6 and HPyV7 lacked a parental strain in the VP regions, whilst avian PyVs lacked parental strains in the LT-Ag regions (Fig. S1). As a consequence, the role played by recombination in shaping the genetic diversity of PyVs remains uncertain.

### Testing the hypothesis of virus–host co-divergence in PyVs

The ParaFit analyses of virus–host divergence yielded comparable results using either the VP1 + 2 or the LT-Ag tree (Table S3). In both tests, the null hypothesis of a random relationship between virus and host trees could not be rejected at the 0.05 level, indicating that the phylogenetic history of PyV is independent of that of the host species from which they were sampled. We also used TreeMap to examine the extent of co-divergence in subtrees that had comparable topologies for VP1 + 2 and LT-Ag. The first pair of subtrees contained KY336, KY157, KY270, KY156, KY397 and the related PyVs. The topologies were identical between VP1 + 2 and LT-Ag (Fig. S8a, b), except for the location of KY270, which had a deeper divergence in the VP1 + 2 tree than in the LT-Ag tree. Importantly, neither dataset exhibited significant virus–host co-divergence (Table S4). The other pair of subtrees contained GTM203, KY369 and their relatives (Fig. S8c, d). Again, no significant signal for co-divergence was detected in these data (Table S4).

## DISCUSSION

We identified and characterized multiple and diverse PyVs circulating in bats, which suggests that bats may play an

important role in PyV evolution and ecology. First, PyVs were widely distributed in the bat populations we tested, being identified in seven of the 35 bat species surveyed, six of which possessed a PyV-positive rate of 10% or above at a given location. Secondly, the genetic diversity of bat PyVs was substantial: in the PyV phylogenies, bat-associated lineages were present in all major PyV clades with the exception of the avian clade. Given that some 1240 bat species have been documented, it is likely that additional biodiversity of PyVs in bats will be observed in the future. Thirdly, the bat PyVs fell in diverse positions across the phylogenetic tree (i.e. they formed a paraphyletic group), indicating that there have been multiple transfers of PyVs among bats and other mammalian species. Finally, because the majority of the specimens were obtained from apparently healthy bats, it is possible that the newly discovered bat PyVs do not cause severe disease in their hosts, although this clearly requires further study.

The identification of distinctive PyVs in bats has greatly expanded our knowledge of the host range of PyVs and in doing so has shed new light on the evolution of these viruses. Instead of using a full-genome tree, we performed statistical tests separately for early- and late-region trees, excluding viruses with highly divergent VP1 + 2 or LT-Ag regions, and hence removing major sources of phylogenetic error. In both our ParaFit and TreeMap analyses, we found no statistically significant evidence for phylogenetic co-divergence across the phylogeny as a whole, thereby ruling out this mode of evolution as an explanation for the entire biodiversity of the PyVs. However, although the hypothesis of co-divergence was rejected overall, there were a number of clear resemblances between the virus and host trees, which are indicative of a long-term interaction between host and virus. For example, PyVs from the same mammalian order often grouped relatively closely together. Although this pattern is suggestive of localized virus–host co-divergence, in theory it is also compatible with preferential host switching, in which species jumps occur most frequently between closely related hosts (Charleston & Robertson, 2002). Preferential host switching has been proposed for a number of viruses, including bat rabies viruses (Streicker *et al.*, 2010), primate lentiviruses (Charleston & Robertson, 2002), *Drosophila* sigma viruses (Longdon *et al.*, 2011), coronaviruses (Cui *et al.*, 2007) and hantaviruses (Ramsden *et al.*, 2009). The respective roles of localized virus–host co-divergence and preferential host switching therefore need to be addressed by sampling a wider range of PyVs from more mammalian orders.

Although bats harbour extensive PyV diversity and group with primates, rodents, sea lions and cows, none are closely related to the known human PyV species. Therefore, bats are unlikely to be the direct source of PyV infection in humans. The close relationship between bat PyVs and (non-human) primate PyVs is of great relevance, as they tended to form monophyletic groups in the phylogenetic trees. Three such monophyletic groups were depicted, namely, (i) MCPyV–PtvPyV1 + 2–GggPyV–KY336–KY157–KY270, (ii) ChPyV–KY156–KY397,

and (iii) JCPyV–BKPyV–SA12–SV40–GTM203 (Fig. 2). All three groups were well supported and suggested that PyV has jumped between bats and primates several times during evolutionary history, although this will need to be confirmed with a wider sampling of mammalian taxa. Less clear was the direction of the host-switching events. One scenario is that PyVs commonly jump from bats to primates, as supported by those cases in which the genetic diversity of primate PyVs fell within that of the bat PyVs. Moreover, the order of transmission within the cluster MCPyV–PtvPyV1 + 2–GggPyV–KY336–KY157–KY270 appeared to be bats to non-human primates and then to humans. However, in other cases, the PyVs of the two mammalian orders formed sister clusters, which is less informative for revealing the direction of host switching. Indeed, the overall diversity of primate PyVs surpassed that of bat PyVs, suggesting that primates are also important PyV reservoirs, although this picture is likely to be strongly influenced by sampling bias.

We established a simple approach to reveal rate heterogeneity in multigene datasets and demonstrated that it occurred frequently within members of the family *Polyomaviridae*. However, the precise causes of this rate heterogeneity remain inconclusive. One possible explanation is intergene recombination with a hotspot between VP1 + 2 and LT-Ag. However, recombination has been reported only rarely in PyVs. For example, although there is sporadic evidence for recombination in JCPyV, it is unclear whether the mosaic sequences in question occur in nature or represent laboratory artefacts (Hatwell & Sharp, 2000; Jobes *et al.*, 1998). In addition, for each of the PyV species that had a signal of recombination, only one parental strain could be identified. As such, the occurrence and importance of recombination in members of the family *Polyomaviridae* remains unclear. Hence, it is equally possible (particularly given the highly divergent nature of the sequences in question) that such a phylogenetic pattern is caused by accelerations in evolutionary rate in different genomic regions in specific viral lineages, although the reasons for any rate accelerations are unclear.

Finally, in light of our analyses, we propose a modification of the current classification system for PyVs. Specifically, we suggest that the highly divergent KIPyV, WUPyV, HPyV6 and HPyV7 should not be classified as a genus because their true evolutionary history is different from that depicted in the late-region trees: KIPyV and WUPyV clustered with JCPyV–BKPyV–SA12–SV40 and GTM203, as well as with SqPyV, BPyV and others, whilst HPyV6 and HPyV7 formed an outgroup to all these species. Moreover, as accurate determination of the phylogenetic history of PyV species is clearly not straightforward, we suggest that it is advisable, on the basis of current data, to retain the old single-genus classification system.

## METHODS

**Sample collection and nucleic acid extraction.** A total of 380 bat specimens were collected from Kenya ( $n=195$ ) in 2006 and from

Guatemala ( $n=185$ ) in 2009. The bats were captured using mist and hand nets. Each bat was measured, sexed and identified to the genus or species level where possible. After euthanasia, a complete necropsy was performed on the captured bats in compliance with approved field protocols. Specimens including blood, major organs, and rectal and oral swabs were obtained and stored immediately on dry ice in the field and were later transferred to  $-70\text{ }^{\circ}\text{C}$  storage before further processing. All protocols for animal capture and use were approved by the CDC Animal Institutional Care and Use Committee, the Kenya Wildlife Service (Nairobi, Kenya), Guatemala Wildlife Service and IACUC from Universidad del Valle de Guatemala.

Each faecal and oral swab was suspended in 200  $\mu\text{l}$  PBS, followed by total nucleic acid extraction using a QIAamp Mini Viral Elute kit (Qiagen). The total nucleic acid was eluted in 80  $\mu\text{l}$  nuclease-free water and used for virus detection.

**PyV detection.** The total nucleic acids extracted from faecal ( $n=286$ ) and oral ( $n=94$ ) swabs were screened for the presence of PyV DNA by semi-nested PCR with consensus degenerate primers (pan-PyV primers) targeted at a conserved region within the VP1 gene using Platinum *Taq* (Invitrogen). Positive and negative PCR controls containing standardized viral DNA extracts and nuclease-free water, respectively, were included in each run. The positive bands of the expected size were purified using a QIAquick Gel Extraction kit (Qiagen). Purified PCR amplicons were sequenced with the PCR primers in both directions on an ABI Prism 3130 Automated Capillary Sequencer (Applied Biosystems).

**Bat mitochondrial gene sequencing.** We amplified and sequenced a cytochrome oxidase subunit I gene fragment (658 bp) to confirm the host species for each of the PyV-positive specimens using a protocol and primers described previously (Hebert *et al.*, 2003).

**Complete genome sequencing.** Twelve representatives of PyV in bats were selected for full-genome sequencing. Sequencing of five representatives failed due to low virus load and significant sequence divergence from the consensus degenerate primers. We applied two strategies: direct PCR amplicon and rolling-circle amplification. The rolling-circle amplification was performed on samples that had a low yield of the PCR amplicon using a TempliPhi 100 Amplification kit (Amersham Biosciences) using a modified protocol (Johns *et al.*, 2006). The initial PCR primer sets for PCR were designed from each pan-PyV amplicon sequence and/or from a conserved region in PyV ST-Ag. As required, walking primers were designed for further PCR and sequencing. The PCRs were performed with an AccuPrime *Taq* DNA Polymerase High Fidelity kit (Invitrogen). All primer sets used in this study are available upon request.

**Phylogenetic analysis.** To determine the phylogenetic relationships of bat PyVs within the family *Polyomaviridae*, we compiled two datasets: (i) a VP1 dataset, which contained all VP1 gene sequences [including 17 partial (270–434 bp) and seven complete sequences] obtained in this study plus representative complete VP1 gene sequences downloaded from GenBank; and (ii) a full-genome dataset, which contained only full genomes sequenced in this study and representative full genomes downloaded from GenBank. For full-genome datasets, only protein-coding regions were used for phylogenetic analyses because the non-coding regions are so divergent among PyV species that they cannot be aligned. The full-genome dataset was further divided by coding regions into a concatenated VP1 + 2 complete gene dataset and a LT-Ag complete gene dataset, which were then analysed separately. This division was based on the observation that the evolutionary histories of these two regions are significantly different (Allander *et al.*, 2007; Gaynor *et al.*, 2007; Krumbholz *et al.*, 2009).

For each dataset (VP1, VP1 + 2 and LT-Ag), taxa that experienced a significant divergence in that gene (or region) were removed to avoid

biasing the phylogenetic inference. Subsequent sequence alignment was performed using MUSCLE v3.6 (Edgar, 2004), using amino acids as a guide for the nucleotide sequence alignment. All ambiguously aligned regions were then removed using Gblocks (Talavera & Castresana, 2007), whilst all third-codon positions were removed to avoid including highly saturated sites in the analysis. Phylogenetic trees were then inferred using MrBayes v3.2 (Ronquist & Huelsenbeck, 2003), assuming a general-time-reversible model of nucleotide substitution with four categories of gamma-distributed rate heterogeneity and a proportion of invariant sites (GTR +  $\Gamma_4$  + I). We used two simultaneous runs of Markov chain Monte Carlo, sampling over 5 million generations, with trees sampled every 1000 generations. The runs were terminated upon convergence (SD of the split frequencies <0.01). The final tree was summarized from both runs with an initial 10% of samples discarded as burn-in.

**Analysis of rate heterogeneity.** We compared pairwise genetic distances to detect variation in the evolutionary rate among PyV genomes. Genetic (patristic) distances were calculated from the MrBayes tree using PATRISTIC v1.0 (Fourment & Gibbs, 2006), or simply as mean percentage divergences. A genetic distance matrix was then derived for each gene alignment. For each pair of these matrices, we calculated a correlation coefficient ( $r$ ) to detect rate heterogeneity, in which a low  $r$  value indicates one or more PyV species with an atypical (usually faster) evolutionary rate in a specific gene. To determine which PyV species experienced a significantly higher evolutionary rate, we plotted genetic distance matrices against each other. Without rate heterogeneity, we would expect a linear relationship between the two matrices. In contrast, comparisons involving rate heterogeneity would deviate significantly from this linear relationship. To verify these results, those PyV species with rate heterogeneity were removed from the dataset and the correlation indexes were then recalculated.

**Analysis of recombination.** The extent of recombination among the PyV species was examined using RDP v3.44 (Martin *et al.*, 2010). We first concatenated gene sequence alignments comprising 39 PyV species examined in this study in the order VP2, VP1 and reverse-complemented LT-Ag. The concatenated alignment was then analysed using the RDP, GENECONV, CHIMAERA, MAXCHI, BOOTSCAN, SISCAN and 3SEQ methods with default parameter settings. All putative recombination events were validated by inferring phylogenetic trees for the recombinant and non-recombinant regions and confirming that they were indeed incongruent.

**Analysis of host–virus co-divergence.** We analysed the extent of co-divergence between virus and host across the entire PyV tree using ParaFit (Legendre *et al.*, 2002), as implemented in the COPYCAT software package (Meier-Kolthoff *et al.*, 2007), for which we calculated pairwise patristic distance matrices. The host genetic distance matrices were derived from the cytochrome oxidase subunit I gene tree, whilst the virus genetic distance matrices were derived from the VP1 + 2 and LT-Ag trees. Each PyV species was represented by a single sequence, and we excluded those PyV species shown to be evolving significantly rapidly. We then calculated the ParaFitGlobal statistic given the distance matrices and host–parasite links. The significance of the test was derived from 99 999 randomizations of the association matrix.

To examine localized host–virus co-divergence, we used the program TreeMap v3.0 (<http://sydney.edu.au/engineering/it/~mcharles/>). Because TreeMap is computationally intensive and part of the virus phylogeny is not well resolved, we limited our analyses to a subset of data that contained fewer than 14 taxa and had relatively consistent topologies for the LT-Ag and VP1 + 2 genes. TreeMap fits the virus topology onto host topology and explains the virus distribution on the host tree through combinations of four types of event: co-divergent (CE), host switches, duplications and losses. The latter three events are collectively

known as non-co-divergent events (NCE). The program calculates optimal combinations of these four events using the Jungles algorithm (Charleston, 1998). To obtain the significance, we performed 100 randomizations of the virus tree, and mapped these trees to the same host tree. Based on these mapping results, we obtained the significance for both CE (percentage of randomizations with equal or more CEs) and NCE (percentage of randomizations with equal or fewer NCEs) with a  $P$  value threshold of 0.05.

## ACKNOWLEDGEMENTS

The study was supported in part by the Global Disease Detection Program TSC funds in the Centers for Disease Control and Prevention, Atlanta, Georgia. We thank Shur-Wern Chern (Division of Viral Diseases, CDC, Atlanta, GA, USA) for her work in the validation of the pan-PyV PCR assay. We thank Ramon Medrano and Maria Luisa Müller from Universidad del Valle de Guatemala (Guatemala City, Guatemala) for support in the fieldwork and thank the Ministry of Agriculture and Ministry of Health from Guatemala. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

## REFERENCES

- Allander, T., Andreasson, K., Gupta, S., Bjerkner, A., Bogdanovic, G., Persson, M. A., Dalianis, T., Ramqvist, T. & Andersson, B. (2007). Identification of a third human polyomavirus. *J Virol* **81**, 4130–4136.
- Bergsten, J. (2005). A review of long-branch attraction. *Cladistics* **21**, 163–193.
- Cantalupo, P., Doering, A., Sullivan, C. S., Pal, A., Peden, K. W., Lewis, A. M. & Pipas, J. M. (2005). Complete nucleotide sequence of polyomavirus SA12. *J Virol* **79**, 13094–13104.
- Charleston, M. A. (1998). Jungles: a new solution to the host/parasite phylogeny reconciliation problem. *Math Biosci* **149**, 191–223.
- Charleston, M. A. & Robertson, D. L. (2002). Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Syst Biol* **51**, 528–535.
- Cui, J., Han, N., Streicker, D., Li, G., Tang, X., Shi, Z., Hu, Z., Zhao, G., Fontanet, A. & other authors (2007). Evolutionary relationships between bat coronaviruses and their hosts. *Emerg Infect Dis* **13**, 1526–1532.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797.
- Fagrouch, Z., Sarwari, R., Lavergne, A., Delaval, M., de Thoisy, B., Lacoste, V. & Verschoor, E. J. (2012). Novel polyomaviruses in South American bats and their relationship to other members of the family *Polyomaviridae*. *J Gen Virol* **93**, 2652–2657.
- Fourment, M. & Gibbs, M. J. (2006). PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. *BMC Evol Biol* **6**, 1.
- Gaynor, A. M., Nissen, M. D., Whiley, D. M., Mackay, I. M., Lambert, S. B., Wu, G., Brennan, D. C., Storch, G. A., Sloots, T. P. & Wang, D. (2007). Identification of a novel polyomavirus from patients with acute respiratory tract infections. *PLoS Pathog* **3**, e64.
- Hatwell, J. N. & Sharp, P. M. (2000). Evolution of human polyomavirus JC. *J Gen Virol* **81**, 1191–1200.
- Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proc Biol Sci* **270**, 313–321.
- Jiang, M., Abend, J. R., Johnson, S. F. & Imperiale, M. J. (2009). The role of polyomaviruses in human disease. *Virology* **384**, 266–273.

- Jobs, D. V., Chima, S. C., Ryschkewitsch, C. F. & Stoner, G. L. (1998). Phylogenetic analysis of 22 complete genomes of the human polyomavirus JC virus. *J Gen Virol* **79**, 2491–2498.
- Johne, R., Wittig, W., Fernández-de-Luaco, D., Höfle, U. & Müller, H. (2006). Characterization of two novel polyomaviruses of birds by using multiply primed rolling-circle amplification of their genomes. *J Virol* **80**, 3523–3531.
- Johne, R., Buck, C. B., Allander, T., Atwood, W. J., Garcea, R. L., Imperiale, M. J., Major, E. O., Ramqvist, T. & Norkin, L. C. (2011). Taxonomical developments in the family *Polyomaviridae*. *Arch Virol* **156**, 1627–1634.
- Khalili, K., White, M. K., Sawa, H., Nagashima, K. & Safak, M. (2005). The agnoprotein of polyomaviruses: a multifunctional auxiliary protein. *J Cell Physiol* **204**, 1–7.
- Krumbholz, A., Bininda-Emonds, O. R., Wutzler, P. & Zell, R. (2009). Phylogenetics, evolution, and medical importance of polyomaviruses. *Infect Genet Evol* **9**, 784–799.
- Kuzmin, I. V., Bozick, B., Guagliardo, S. A., Kunkel, R., Shak, J. R., Tong, S. & Rupprecht, C. E. (2011). Bats, emerging infectious diseases, and the rabies paradigm revisited. *Emerging Health Threats Journal* **4**, 7159.
- Legendre, P., Desdevises, Y. & Bazin, E. (2002). A statistical test for host–parasite coevolution. *Syst Biol* **51**, 217–234.
- Longdon, B., Hadfield, J. D., Webster, C. L., Obbard, D. J. & Jiggins, F. M. (2011). Host phylogeny determines viral persistence and replication in novel hosts. *PLoS Pathog* **7**, e1002260.
- Martin, D. P., Lemey, P., Lott, M., Moulton, V., Posada, D. & Lefevre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462–2463.
- Meier-Kolthoff, J. P., Auch, A. F., Huson, D. H. & Göker, M. (2007). COPYCAT: cophylogenetic analysis tool. *Bioinformatics* **23**, 898–900.
- Misra, V., Dumonceaux, T., Dubois, J., Willis, C., Nadin-Davis, S., Severini, A., Wandeler, A., Lindsay, R. & Artsob, H. (2009). Detection of polyoma and corona viruses in bats of Canada. *J Gen Virol* **90**, 2015–2022.
- Norkin, L. C., Allander, T., Atwood, W. J., Buck, C. B., Garcea, R. L., Imperiale, M. J., Johne, R., Major, E. O., Pipas, J. M. & Ramqvist, T. (2011). *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Oxford: Elsevier.
- Pérez-Losada, M., Christensen, R. G., McClellan, D. A., Adams, B. J., Viscidi, R. P., Demma, J. C. & Crandall, K. A. (2006). Comparing phylogenetic codivergence between polyomaviruses and their hosts. *J Virol* **80**, 5663–5669.
- Pipas, J. M. (1992). Common and unique features of T antigens encoded by the polyomavirus group. *J Virol* **66**, 3979–3985.
- Ramsden, C., Holmes, E. C. & Charleston, M. A. (2009). Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Mol Biol Evol* **26**, 143–153.
- Renshaw, R. W., Wise, A. G., Maes, R. K. & Dubovi, E. J. (2012). Complete genome sequence of a polyomavirus isolated from horses. *J Virol* **86**, 8903.
- Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
- Schowalter, R. M., Pastrana, D. V., Pumphrey, K. A., Moyer, A. L. & Buck, C. B. (2010). Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. *Cell Host Microbe* **7**, 509–515.
- Shadan, F. F. & Villarreal, L. P. (1993). Coevolution of persistently infecting small DNA viruses and their hosts linked to host-interactive regulatory domains. *Proc Natl Acad Sci U S A* **90**, 4117–4121.
- Siebrasse, E. A., Reyes, A., Lim, E. S., Zhao, G., Mkakosya, R. S., Manary, M. J., Gordon, J. I. & Wang, D. (2012). Identification of MW polyomavirus, a novel polyomavirus in human stool. *J Virol* **86**, 10321–10326.
- Soeda, E., Maruyama, T., Arrand, J. R. & Griffin, B. E. (1980). Host-dependent evolution of three papova viruses. *Nature* **285**, 165–167.
- Streicker, D. G., Turmelle, A. S., Vonhof, M. J., Kuzmin, I. V., McCracken, G. F. & Rupprecht, C. E. (2010). Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* **329**, 676–679.
- Talavera, G. & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564–577.
- Tong, S., Conrardy, C., Ruone, S., Kuzmin, I. V., Guo, X., Tao, Y., Niezgodna, M., Haynes, L., Agwanda, B. & other authors (2009). Detection of novel SARS-like and other coronaviruses in bats from Kenya. *Emerg Infect Dis* **15**, 482–485.
- Tong, S., Li, Y., Rivailier, P., Conrardy, C., Castillo, D. A., Chen, L. M., Recuenco, S., Ellison, J. A., Davis, C. T. & other authors (2012). A distinct lineage of influenza A virus from bats. *Proc Natl Acad Sci U S A* **109**, 4269–4274.
- Van Ghelue, M., Khan, M. T., Ehlers, B. & Moens, U. (2012). Genome analysis of the new human polyomaviruses. *Rev Med Virol* **22**, 354–377.
- White, M. K., Safak, M. & Khalili, K. (2009). Regulation of gene expression in primate polyomaviruses. *J Virol* **83**, 10846–10856.