AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology

Assessing Host-Virus Codivergence for Close Relatives of Merkel Cell Polyomavirus Infecting African Great Apes

Nadège F. Madinda,[a,b,c] Bernhard Ehlers,[d] Joel O. Wertheim,[e] Chantal Akoua-Koffi,[f] Richard A. Bergl,[g] Christophe Boesch,[b] Dieudonné Boji Mungu Akonkwa,[h] Winnie Eckardt,[i,j] Barbara Fruth,[k,l] Thomas R. Gillespie,[j,m] Maryke Gray,[n] Gottfried Hohmann,[b] Stomy Karhemere,[o] Deo Kujirakwinja,[p] Kevin Langergraber,[b,q] Jean-Jacques Muyembe,[o] Radar Nishuli,[h] Maude Pauly,[a,d]* Klara J. Petrzelkova,[r,s,t,u] Martha M. Robbins,[b] Angelique Todd,[v] Grit Schubert,[a] Tara S. Stoinski,[i,j] Roman M. Wittig,[b,w] Klaus Zuberbühler,[x,y,z] Martine Peeters,[aa,bb] Fabian H. Leendertz,[a] Sébastien Calvignac-Spencer[a]

Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany[a]; Department of Primatology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany[b]; Institut de Recherches en Ecologie Tropicale, Libreville, Gabon[c]; FG12, Measles, Mumps, Rubella and Viruses Affecting Immunocompromised Patients, Robert Koch Institute, Berlin, Germany[d]; Department of Medicine, University of California, San Diego, California, USA[e]; Centre de Recherche pour le Développement, Université Alassane Ouattara de Bouake, Bouake, Côte d'Ivoire[f]; North Carolina Zoological Park, Asheboro, North Carolina, USA[g]; Institut Congolais pour la Conservation de la Nature, Kinshasa, Democratic Republic of Congo[h]; Diane Fossey Gorilla Fund International, Atlanta, Georgia, USA[i]; Department of Environmental Sciences and Program in Population Biology, Ecology and Evolution, Emory University, Druid Hills, Georgia, USA[j]; Division of Neurobiology, Ludwig Maximilians University, Munich, Germany[k]; Centre for Research and Conservation, Royal Zoological Society of Antwerp, Antwerp, Belgium[l]; Department of Environmental Health, Rollins School of Public Health, Emory University, Druid Hills, Georgia, USA[m]; International Gorilla Conservation Program, Kigali, Rwanda[n]; Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of Congo[o]; Wildlife Conservation Society, Grauer's Gorilla Project, Kinshasa-Gombe, Democratic Republic of Congo[p]; School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA[q]; Institute of Vertebrate Biology, Academy of Sciences, Brno, Czech Republic[r]; Department of Pathology and Parasitology, University of Veterinary and Pharmaceutical Sciences, Brno, Czech Republic[s]; Biology Centre, Institute of Parasitology, Academy of Sciences of the Czech Republic, Ceske Budejovice, Czech Republic[t]; Liberec Zoo, Liberec, Czech Republic[u]; World Wildlife Foundation, Dzanga Sangha Protected Areas, Bangui, Central African Republic[v]; Taï Chimpanzee Project, Centre Suisse de Recherches Scientifiques, Abidjan, Côte d'Ivoire[w]; Institute of Biology, University of Neuchatel, Neuchatel, Switzerland[x]; Budongo Conservation Field Station, Masindi, Uganda[y]; School of Psychology, University of St. Andrews, St. Andrews, Scotland, United Kingdom[z]; Unité Mixte Internationale 233, Institut de Recherche pour le Développement, INSERM U1175, and University of Montpellier, Montpellier, France[aa]; Computational Biology Institute, Montpellier, France[bb]

ABSTRACT

It has long been hypothesized that polyomaviruses (PyV; family *Polyomaviridae*) codiverged with their animal hosts. In contrast, recent analyses suggested that codivergence may only marginally influence the evolution of PyV. We reassess this question by focusing on a single lineage of PyV infecting hominine hosts, the Merkel cell polyomavirus (MCPyV) lineage. By characterizing the genetic diversity of these viruses in seven African great ape taxa, we show that they exhibit very strong host specificity. Reconciliation analyses identify more codivergence than noncodivergence events. In addition, we find that a number of host and PyV divergence events are synchronous. Collectively, our results support codivergence as the dominant process at play during the evolution of the MCPyV lineage. More generally, our results add to the growing body of evidence suggesting an ancient and stable association of PyV and their animal hosts.

IMPORTANCE

The processes involved in viral evolution and the interaction of viruses with their hosts are of great scientific interest and public health relevance. It has long been thought that the genetic diversity of double-stranded DNA viruses was generated over long periods of time, similar to typical host evolutionary timescales. This was also hypothesized for polyomaviruses (family *Polyomaviridae*), a group comprising several human pathogens, but this remains a point of controversy. Here, we investigate this question by focusing on a single lineage of polyomaviruses that infect both humans and their closest relatives, the African great apes. We show that these viruses exhibit considerable host specificity and that their evolution largely mirrors that of their hosts, suggesting that codivergence with their hosts played a major role in their diversification. Our results provide statistical evidence in favor of an association of polyomaviruses and their hosts over millions of years.

Viral diversification is notably shaped by processes that promote host specificity, for example, antagonistic coevolution (1), and opportunities to colonize new hosts, i.e., cross-species transmission events. Depending on their balance, host-virus codivergence patterns may arise and persist over the long term. Long-term codivergence may have played an important role in the diversification of some double-stranded DNA (dsDNA) viruses, e.g., herpesviruses and papillomaviruses (2–5).

Polyomaviruses (PyV; family *Polyomaviridae*) are small nonenveloped viruses with a circular double-stranded DNA genome (ca. 5 kb in length) (6). They infect a broad range of animals, including arthropods and vertebrates (fish, birds, and mammals), and comprise at least 13 distinct viruses infecting humans (7, 8). In humans, infections occur in childhood, persist lifelong, and are usually asymptomatic (9). At least five PyV have been associated with disease in immunosuppressed individuals (10–12). Routes of transmission are poorly characterized but may involve respiratory droplets and/or environmental contamination.

Putative codivergence events of hosts and their PyV have repeatedly been invoked in the literature to explain the structure of PyV diversity. Reconciliation analyses performed on the family scale sometimes supported a significant contribution of codivergence events (8, 13), but other studies have failed to detect any global codivergence signal (14, 15). Similarly, authors focusing on

more recent evolutionary events defended opposing views as to the potential codivergence of humans and JC polyomaviruses (JCV) (16, 17–19). An alternative scenario combining ancient noncodivergence events and subsequent lineage-specific codivergence with their hosts, as proposed for papillomaviruses (3), still remains to be tested. The disparate sampling of the PyV animal hosts as well as the lack of resolution of many internal branches of this viral family tree severely compromises the power to detect such patterns from currently available data.

To overcome these limitations, we designed a formal test to assess the influence of codivergence on the evolution of PyV and characterized the genetic diversity of a single lineage of PyV that infects a set of recently diverged host species with a well-resolved phylogeny. Specifically, we focused on viruses infecting African great apes (here, referred to simply as great apes) belonging to the lineage comprising the Merkel cell polyomavirus (MCPyV), an oncogenic human virus (*Human polyomavirus 5*, genus *Alphapolyomavirus*) (10, 20, 21, 22).

MATERIALS AND METHODS

Samples. We collected a total of 386 fecal samples in the wild from seven great ape taxa (Table 1). Great ape samples were collected opportunistically or from habituated animals and preserved in RNAlater (Qiagen, Hilden, Germany), in liquid nitrogen, or by drying over silica. We also collected 197 fecal samples from two human populations in Côte d'Ivoire and the Democratic Republic of the Congo (Table 1). Human samples were preserved in liquid nitrogen. For animal samples, authorization was obtained from responsible local authorities. For human samples, institutional authorization was received along with the written consent of all participants in the study.

Molecular biology. DNA extraction was performed using a Roboklon stool kit (Roboklon, Berlin, Germany), according to manufacturer's instructions.

To identify Merkel cell polyomavirus (MCPyV)-related sequences in DNA extracts, a nested PCR assay was set up that made use of generic, degenerate primers targeting a ca. 700-bp VP1 fragment (Table 2, PCR1). These primers were designed on the basis of published MCPyV sequences and those of MCPyV-related PyV of nonhuman primates (NHP). First-round PCR mixes were set up so as to reduce the risk of carryover contamination with PCR products. The mixtures contained 0.2 μM each primer, 200 μM deoxynucleoside triphosphate (dNTP) mix (with dUTP replacing dTTP), 0.3 U of AmpErase uracil *N*-glycosylase (UNG; Invitrogen, Carlsbad, CA, USA), 4 mM $MgCl_2$, 1× PCR buffer, and 1.25 U of Platinum Taq polymerase (Invitrogen). Second-round PCR mixes were prepared in the same way but did not include UNG. Cycling conditions were as follows: 7 min at 45°C (UNG activity) and 7 min at 95°C, followed by 47 cycles (first round) or 45 cycles (second round) of 30 s at 95°C, 30 s at 57°C (first round) or 58°C (second round), and 2 min at 72°C, with a final 10 min at 72°C.

Twenty-two positive samples were then selected based on the results of preliminary phylogenetic analyses to attempt additional nested long-distance (LD) amplification of partial genomes (approximately 2.5 kb) with generic, degenerate primers (Table 2, PCR2) using a TaKaRa-Ex kit (TaKaRa Bio, Inc., Otsu, Japan) according to the manufacturer's instructions. Nondegenerate primers (sequences available from the authors upon request) were used for amplification of the remaining part (approximately 2.8 kb) of the genome with LD nested PCR. LD PCR cycling conditions followed those reported in Leendertz et al. (21).

A total of 174 human DNA extracts were also screened using a semi-nested PCR system targeting a ca. 200-bp VP1 fragment (Table 2, PCR3). This system was designed to be specific to members of lineage 1 (see below) and was validated on a selection of great ape DNA extracts of known status before being employed on human DNA extracts (data not shown). PCR mix preparation and cycling conditions followed those mentioned above.

Short PCR products were purified using ExoSAP-IT (Affymetrix, Santa Clara, CA, USA) whereas LD PCR products were purified using a column-based PCR purification kit (Qiagen, Venlo, Netherlands). All purified products were sequenced with a BigDye Terminator cycle sequencing kit on a 377 DNA automated sequencer (Applied Biosystems, Warrington, United Kingdom).

Overlapping partial sequences were used to reconstruct circular genomes using Geneious, version 7.1.4 (Biomatters, Ltd., Auckland, New Zealand) (23). Genomes were subsequently annotated with Geneious.

Phylogenetic analyses. Partial VP1 and complete genome data sets were assembled that comprised sequences generated in this study and a selection of (partial VP1) or all (complete genome) MCPyV sequences as well as any publicly available great ape MCPyV-related sequences. Both data sets were reduced to unique sequences and aligned using MUSCLE, as implemented in SeaView, version 4 (24). Conserved nucleotide blocks were selected from the alignments using Gblocks (in SeaView) (25) and used for recombination analyses using RDP4, version 4.46 (26). Two final alignments were generated: one with 74 sequences and 838 positions and another with 16 sequences and 5,150 positions. Further analyses were performed only on the partial VP1 alignment as this comprised more genetic diversity.

The best model of nucleotide substitution (general time-reversible matrix with rate variation across sites [$GTR+G_4$]) was selected with jModelTest, version 2.1.4 (27), using the Bayesian information criterion. Maximum likelihood (ML) analyses were performed using PhyML, version 3 (28), as implemented on the PhyML Web server (29). The best-fit root of the ML tree was identified using TempEst, version 1.5 (30; http://tree.bio.ed.ac.uk/software/tempest/). Bayesian Markov chain Monte Carlo (BMCMC) analyses were performed in BEAST, version 1.8.2, under a log-normal relaxed clock (uncorrelated) and three different models of diversification: a pure coalescent model assuming a constant population size, a multispecies coalescent model using the 14-species scheme suggested by species delineation analyses (see below), and a birth-death speciation model (31, 32). Convergence of BMCMC runs (at least two runs per model) and appropriate sampling of the posterior were checked with Tracer, version 1.6 (http://tree.bio.ed.ac.uk/software/tracer/). Branch robustness was assessed through nonparametric bootstrapping (250 pseudoreplicates for ML analyses) or posterior probabilities (BMCMC analyses).

Host specificity analyses. Host specificity was assessed by running Bayesian tip-association significance testing (BaTS) on all posterior samples of trees (PST) generated by BMCMC analyses (33). BaTS allows for tests of the correlation of trait states with ancestry while accounting for phylogenetic uncertainty suggested by the PST. It compares observations to a null distribution generated under the assumption that trait values are not influenced by ancestry. Host species/subspecies was defined as the

Received 22 February 2016 **Accepted** 12 July 2016

Accepted manuscript posted online 20 July 2016

Citation Madinda NF, Ehlers B, Wertheim JO, Akoua-Koffi C, Bergl RA, Boesch C, Akonkwa DBM, Eckardt W, Fruth B, Gillespie TR, Gray M, Hohmann G, Karhemere S, Kujirakwinja D, Langergraber K, Muyembe J-J, Nishuli R, Pauly M, Petrzelkova KJ, Robbins MM, Todd A, Schubert G, Stoinski TS, Wittig RM, Zuberbühler K, Peeters M, Leendertz FH, Calvignac-Spencer S. 2016. Assessing host-virus codivergence for close relatives of Merkel cell polyomavirus infecting African great apes. J Virol 90:8531–8541. doi:10.1128/JVI.00247-16.

Editor: S. R. Ross, University of Illinois at Chicago

Address correspondence to Sébastien Calvignac-Spencer, calvignacs@rki.de.

* Present address: Maude Pauly, Department of Infection and Immunity, Luxembourg Institute of Health, Esch-sur-Alzette, Luxembourg.

TABLE 1 Samples and screening results

Species and subspecies	Country	Site	No. of samples	No. of positives	Proportion (% [95% CI])[a]	Minimum identity within host subspecies (%)[a]	Maximum identity with a publicly available sequence (% [GenBank accession no. and host subspecies])[b]
Gorilla gorilla gorilla	Cameroon	Belgique	20	1			
		Campo Ma'an National Park	18	0			
		Mambele	19	0			
	Central African Republic	Dzanga-Sangha Special Reserve	23	0			
	Gabon	Loango National Park	25	1	1.9 (0.3–7.4)	98.5	99 (HQ385752, *Gorilla gorilla gorilla*)
Gorilla gorilla diehli	Cameroon	Greater Takamanda-Mone Landscape	22	0	0 (0–18.5)	NA	NA
Gorilla beringei beringei	Rwanda	Volcanoes National Park	51	0			
	Uganda	Bwindi Impenetrable National Park	30	1	1.2 (0–7.6)	NA	98 (HQ385752, *Gorilla gorilla gorilla*)
Gorilla beringei graueri	Democratic Republic of the Congo	Kahuzi-Biega National Park	34	7	20.6 (9.3–38.4)	74.7	99 (HQ385752, *Gorilla gorilla gorilla*)
							83 (HQ385747, *Pan troglodytes verus*)
Pan troglodytes troglodytes	Cameroon	Belgique	5	1			
		Cameroun	10	1			
		Campo Ma'an National Park	1	0			
		Mambele	9	1			
	Gabon	Loango National Park	27	3	11.5 (4.8–24.1)	77	95 (HQ385748, *Pan troglodytes verus*)
							94 (HQ385747, *Pan troglodytes verus*)
Pan troglodytes schweinfurthii	Uganda	Budongo Central Forest Reserve	33	9			
		Kibale Forest National Park	33	11	30.3 (20–43)	76.5	95 (HQ385747, *Pan troglodytes verus*)
							94 (HQ385748, *Pan troglodytes verus*)
Pan paniscus	Democratic Republic of the Congo	Salonga National Park	26	14	53.8 (33.7–72.9)	77.4	91 (HQ385751, *Pan troglodytes verus*)
							91 (HQ385746, *Pan troglodytes verus*)
Homo sapiens	Côte d'Ivoire	Taï National Park	96	16			
	Democratic Republic of the Congo	Salonga National Park	101	14	15.2 (10.7–21.2)	99	100 (JF812999, *Homo sapiens*)

[a] At the species/subspecies level. NA, not assessed; CI, confidence interval.

[b] PyV sequences from Western chimpanzees (*Pan troglodytes verus*) were already available from a previous study (14).

TABLE 2 Primers used in this study

PCR	Primer name	Primer sequence (5′–3′)	Annealing temp (°C)	Fragment size (kb)
PCR1	PCR1.1-f	TGTGCTCCTAAGCCBGGATG	57	
	PCR1.1-r	ACTACTGGGTATGGRTTYTTMACC		
	PCR1.2-f	CTGAATCCAAGRATGGGAGT	58	0.7
	PCR1.2-r	CATGAAANGCCATTTTNCCACT		
PCR2	PCR2.1-f	CTGAAGYCTGGGACGMTGAG	57	
	PCR2.1-r	GCAAACATRTGRTAATTGACTCCC		
	PCR2.2-f	TCAGACWCCSAGTCCAGAGG	58	2.5
	PCR2.2-r	GCAAATCYARRGGYTCTCCTC		
PCR3	PCR3.1-f	TGATATGCAGCCMAATMWWCARC	58	
	PCR3.1-r	AAACATGTGATAATTGACTCCCTC		
	PCR3.1-f	TGATATGCAGCCMAATMWWCARC	58	0.2
	PCR3.2-r	AATTGACTCCCTCAATAGGAATG		

trait of interest. Its association with ancestry was assessed at the host subspecies level (8 states) and species level (5 states) independently by running separate BaTS analyses during which 500 null replicates per tree were generated. Global as well as state-specific statistics of association were computed (global, association index [AI] and Fitch parsimony score [PS]; state-specific, maximum exclusive single-state clade size [MC]).

To investigate the association of host and PyV diversification processes, we performed PyV species delineation analyses with the R package SPLITS (34), using the maximum clade credibility tree derived from BMCMC analyses performed under the (coalescent) constant population size model. SPLITS implements general mixed Yule coalescent (GMYC) models (34, 35) which are optimized and compared to the null hypothesis that the tree was generated by pure coalescent processes, i.e., that it reflects diversity within a single species. When the GMYC model outperforms the null model, the parts of the tree most likely to have been generated by between-species and within-species processes can be identified, thereby delineating species (according to the phylogenetic species concept).

Codivergence analyses. The degree of topological congruence and the number of events necessary to explain (reconcile) incongruences were assessed using Jane, version 4 (36). Jane implements a genetic algorithm to quickly identify the most parsimonious scenarios of coevolution involving several types of events (codivergence, duplication, duplication with host switch, loss, and failure to diverge). As input, it requires host and parasite phylogenies and the respective tip mapping as well as an event cost matrix. A simplified version of the PyV phylogeny was used as input, whereby single-host clades were collapsed. Three sets of costs were tested: (i) set 1 with the parameter codivergence set at 0, duplication at 1 (under the assumption that duplication incurs costs related to within-host speciation, e.g., maintaining of distinct lineages in the face of within-host competition or tropism change within the same host), duplication with host switch at 1 (host switch incurs costs), loss at 1 (prevalence was always high), and failure to diverge at 1 (given their respective evolutionary timescales, viruses are unlikely to fail to diverge when their hosts do so); (ii) set 2, with the same parameters as set 1 but with loss set at 0 (prevalence may have been low at some point in the past); (iii) set 3, with codivergence set at −1 and all noncodivergence events at 0. Set 3 is a variation of set 1 with the same relative costs but where all costs are shifted to the left. This allows costs and codivergence events to be equated. Jane was run using the vertex-based cost mode, and the parameters of the genetic algorithm were kept at their default values (population size, 100; number of generations, 100). To determine the probability of observing the inferred costs by chance, costs were also calculated on a set of 500 samples for which tip mapping was randomized. Settings of the genetic algorithm were kept at default values.

Topology tests were performed to assess whether exceptions to a scenario of perfect codivergence observed in the PyV phylogenetic tree were better supported by the data than a perfect codivergence model. This was done by using approximately unbiased (AU) tests, as implemented in CONSEL, version 0.1i (37).

Finally, divergence dates were also estimated. Topological congruence could emerge independently of codivergence, e.g., through preferential host switching (38). Observing synchronicity in timing of divergence events of hosts and their parasites reinforces the codivergence hypothesis. When viral lineage duplication occurs, synchronicity of parasite divergence events is also expected (provided the viral lineages maintain similar degrees of association to their host). Divergence date estimates were obtained using two methods: (i) as part of the aforementioned BMCMC analyses or (ii) by reestimating branch lengths of the ML tree under codon models using HyPhy, version 2.2.4 (39), and making the resulting tree ultrametric using a relaxed clock model implemented in r8s (40). The codon models used for this second set of analyses were a pure branch model derived of MG94, in which the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site is estimated for each branch but assumed to be unchanged across sites (41), and an adaptive branch site random effects model (aBSREL), in which this ratio is estimated for each branch and allowed to vary across sites (42). We detected marked saturation at synonymous sites (data not shown); such strong saturation complicates analyses under both nucleotide and codon models. For both BMCMC and ML-based analyses, the relaxed clock was calibrated by setting a prior distribution (BEAST) or enforcing a fixed age (r8s) for the time to the most recent common ancestor (tMRCA) of lineage 1 using a published estimate of the split date of all hominine species (either 5.6 million years or a normal distribution with a mean of 5.6 million years and standard deviation of 0.25 million years [43]). Because we used the split date of all hominine species, estimates of times to the most recent common ancestors for viruses should be regarded as minimum bounds (viral coalescence times will necessarily predate the effective ancestral host population/species split). It should also be noted that divergence dates of the different hominine lineages are a point of active debate; this stems from both a scarce paleontological record and uncertainty in estimates of long-term mutation rates at genomic scales. For example, the estimate we opted for here (5.6 million years) is drawn from genomic analyses that proposed two estimates (5.6 or 11.2 million years), depending on priors on the substitution rates (1×10^{-9} or 0.5×10^{-9} mutations $\cdot$ bp^{-1} $\cdot$ year^{-1}) (43). The focus of our synchronicity analyses was, however, on relative internode lengths, not absolute dates. Calendar years can thus be replaced with genetic distances and/or ratios of interest (see Table 6).

Accession numbers. Partial VP1 and whole-genome sequences, respectively, were deposited at the European Nucleotide Archive and

GenBank under accession numbers LT158307 to LT158400 and KT184856 to KT184862.

Data availability. r8s and BEAUTi XML exemplary input files are available from the authors upon request.

RESULTS

Detection of short MCPyV-related sequences. Using a specific PCR system designed to amplify a ca. 700-bp fragment of the VP1 gene, we screened 386 fecal great ape and 197 human samples (Table 1). We detected MCPyV-related sequences in 50 great ape DNA extracts representing all hosts but *Gorilla gorilla diehli*, with fecal detection rates between 1.2% (*Gorilla beringei beringei*) and 53.8% (*Pan paniscus*). Nearly all sequences were found only at one site; a single sequence was detected in five and two Eastern chimpanzees (*Pan troglodytes schweinfurthii*) at two distinct sites in Uganda. For species/subspecies from which more than two sequences were obtained, considerable sequence divergence was observed; e.g., maximum observed distances were over 20%, possibly reflecting the circulation of viruses belonging to different lineages (discussed in more detail below). Minimum observed distances to publicly available sequences were often relatively high, i.e., between 5 and 17%. Finally, we also detected MCPyV sequences with >99% identity to published MCPyV sequences in 30 human DNA extracts (fecal detection rate, 15.2%). Most human DNA extracts were also screened with a PCR system intended to be lineage 1 specific (see below); all assays were negative.

Characterization of full genomes. We attempted to determine full-genome sequences from a selection of DNA extracts ($n = 22$). This was possible for samples from *P. paniscus* ($n = 2$), *P. troglodytes troglodytes* ($n = 3$), *P. troglodytes schweinfurthii* ($n = 1$), and *Gorilla beringei graueri* ($n = 1$). Examination of putative open reading frames (ORFs) showed that all genomes displayed a typical PyV genome structure with an early region encoding regulatory proteins (small T and large T antigens) and a late region coding for structural proteins (VP1, VP2, and VP3) separated by a noncoding control region (NCCR). No open reading frame likely to encode a putative agnoprotein was identified. Overall, a ca. 80% sequence similarity to genomes of MCPyV and MCPyV-related nonhuman primate PyV was observed. Preliminary analyses revealed that the full genomes represented only a fraction of the overall genetic diversity detected in this study. To incorporate this broader diversity, we performed all of following phylogenetic analyses on an alignment of partial VP1 sequences (including sequences extracted from the novel full genomes).

Molecular phylogeny. We could not detect any signal indicative of recombination in the VP1 alignment (26). Phylogenetic analyses in both maximum likelihood (ML) (28) and Bayesian (31) frameworks supported the existence of a number of host-specific clades (Fig. 1 and 2). All clades seemed to derive from three ancient lineages: one that comprised only MCPyV sequences and two that included only viral sequences detected in gorillas, bonobos, and chimpanzees. Branching order partially recapitulated host divergence events in the two great ape lineages (here, referred to as lineages 1 and 2) (Fig. 1 and 2). We identified four exceptions: (i) the polyphylies of PyV infecting Western chimpanzees in lineage 1 and Eastern chimpanzees in lineage 2, (ii) the interspersion of PyV infecting Eastern lowland and mountain gorillas in lineage 1, and (iii) the basal position of MCPyV.

Host specificity. We estimated the statistical support for host specificity using BaTS (Table 3). We found that viral sequences found in a single host species were generally more likely to be closely related than expected by chance when both global and state-specific statistics were considered. The only exceptions corresponded to viral sequences identified in the sister subspecies *G. beringei beringei* and *G. beringei graueri*.

We also characterized the viral diversification process by running a species delineation analysis using general mixed Yule coalescent models (GMYC) (34, 35). The best GMYC model outperformed the null, full coalescent model ($P = 0.0005$) and identified 14 entities, among which 10 comprised several sequences. Nine multisequence entities comprised only sequences identified from a single host species/subspecies, indicating a close parallelism of PyV and host diversification processes (Fig. 1).

Codivergence. Taking the viral phylogeny presented in Fig. 1 as a given, we performed reconciliation analyses using Jane (Table 4). Under all tested cost sets, and whether the host species or subspecies phylogeny was considered, the number of codivergence events always exceeded the number of noncodivergence events. Randomization tests showed that, irrespective of the cost set, these results could not be explained by chance at the subspecies level. At the species level and using a P value threshold of 0.05, results obtained under two of the cost sets failed to reach statistical significance; it should, however, be noted that the species-level phylogeny comprises only five species, meaning that these tests had low power.

We also examined whether the viral topology presented in Fig. 1 was a better fit to our data than alternative topologies which enforced strict codivergence within lineages 1 and 2. The model forcing MCPyV to belong to lineage 1 was the only one that was rejected (AU test; $P = 0.003$). Monophyly of PyV infecting Western chimpanzees in lineage 1 and Eastern chimpanzees in lineage 2 as well as inclusion of MCPyV in lineage 2 could not be excluded (AU test; $P = 0.52$, 0.13, and 0.11, respectively). Given the very recent split of Eastern lowland and mountain gorillas (about 10,000 years ago [44]), the interspersion of PyV infecting these subspecies appeared biologically plausible, so we did not compare this scenario to a strict codivergence model.

Besides topological congruence, codivergence should result in synchronization of (i) viral and host divergence dates and (ii) viral divergence dates in the case of ancestral viral lineage duplication. We first estimated divergence dates using a relaxed clock model applied to nucleotide data in a Bayesian framework. For five of the six focal nodes of our analyses (nodes 1.2 to 4 and 2.1 to 3), these estimates were significantly older than host divergence events (Table 5). This pattern was compatible with the effects of the time dependency of molecular rates, i.e., the decay of molecular rates with increasing observation timescales, which can result in overestimating recent time to the most recent common ancestor (tMRCA) inferred from deep calibration points (19, 45–47). As this may arise through the effects of unaccounted-for purifying selection (among other possible mechanisms) (48, 49, 50), we reestimated all branch lengths using selection-aware models of codon evolution in an ML framework. A branch model of codon evolution resulted in divergence dates very close to those inferred by BMCMC analyses. Using an adaptive branch site random effect model of codon evolution, strong purifying selection was detected on a number of branches, including deep ones (data not shown). Most of the resulting increase in the overall tree length was supported by a single basal branch. This expansion prevented us from deriving any trustworthy tMRCA estimates.

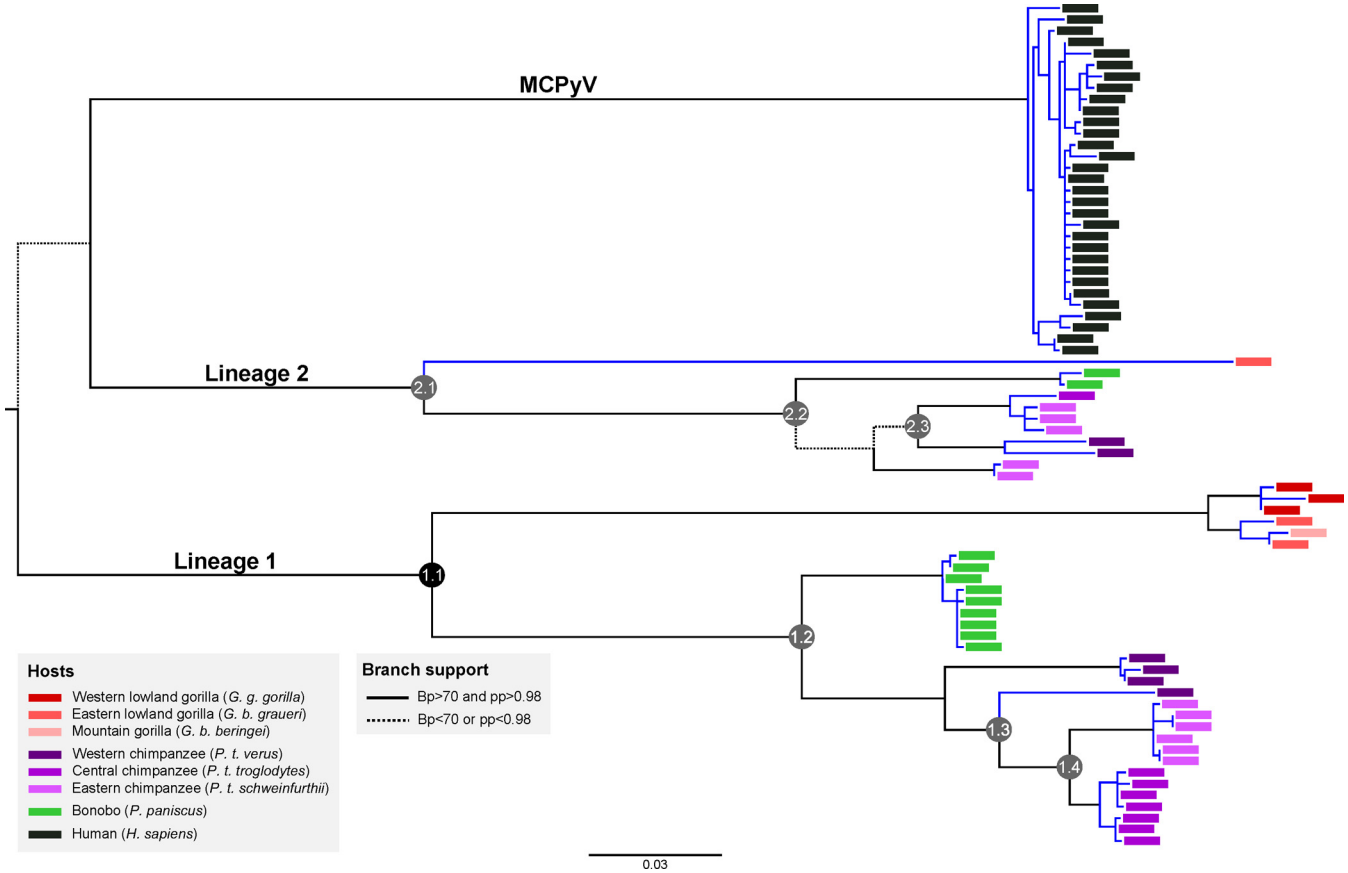

FIG 1 Maximum likelihood tree derived from an alignment of partial VP1 sequences. This tree was rooted at its center. The six gray circles stand for the main nodes whose date estimates are given in full in Tables 5 and 6; the black circle indicates the node that was used to calibrate the analyses. Note that these circles coincide with putative codivergence events. This tree was rooted using TempEst. Bp, bootstrap; pp, posterior probability. *G. g. gorilla*, *G. gorilla gorilla*; *G. b. graueri*, *G. beringei graueri*; *G. b. beringei*, *G. beringei beringei*; *P. t. verus*, *P. troglodytes verus*; *P. t. troglodytes*, *P. troglodytes troglodytes*; *P. t. schweinfurthii*, *P. troglodytes schweinfurthii*; *H. sapiens*, *Homo sapiens*.

Given the likely impact of strong purifying selection and our inability to properly account for it, we reexamined branch length/internode ratios by rescaling the results in Table 5, using the tMRCA of a young node, node 1.4 (divergence of lineage 1 PyV infecting *P. troglodytes troglodytes* and *P. troglodytes schweinfurthii*), as a new arbitrary unit (Table 6). This resulted in a good agreement of host and virus relative divergence dates for most nodes (nodes 1.3 and 2.3 and nodes 1.2 and 2.2). The tMRCA of lineage 2 PyV infecting all great apes was a large underestimate of the divergence date of their hominine hosts, as expected under the hypothesis that deep branch lengths are severely underestimated.

DISCUSSION

The lack of any physical viral fossil record considerably complicates the task of understanding the long-term association of viruses with their hosts. However, using their present-day distribution, their nucleic acid sequences, and (more rarely) other biological traits, we can try to infer how long and how closely viruses have been associated with their hosts. The aim of this study was to determine whether codivergence, i.e., viral diversification driven by host diversification, is an important driver of PyV evolution.

Measurable host specificity is an absolute prerequisite for characterizing historical codivergence events. Host specificity has often been assumed for PyV, with only a few well-identified exceptions, e.g., budgerigar fledgling disease virus and simian virus 40 (SV40). Over the last decade, this assumption has been repeatedly supported by the implementation of generic PyV detection tools which have not revealed any multihost PyV (20, 51). Here, we used a PCR assay designed to specifically target a single PyV lineage to generate a large sample of sequences from closely related PyV infecting wild African great apes. Statistical tests strongly supported marked host specificity, which was still detectable at the host subspecies level. Viral diversification/speciation—as revealed by a GMYC model, i.e., according to the phylogenetic species concept—appeared strongly influenced by host diversification.

Host specificity and a coupling of viral diversification/speciation with host diversity could also arise over much shorter timescales than those implied by codivergence events. If codivergence is a dominant evolutionary process, a key expectation is that virus and host phylogenies should often be congruent. Phylogenetic analyses of great ape MCPyV-like sequences highlighted the existence of two viral lineages within which viral divergence events were mostly in line with hominine divergence events. Exceptions to the expectation of perfect codivergence within these lineages were not statistically supported. In addition, reconciliation analyses identified more codivergence events than noncodivergence

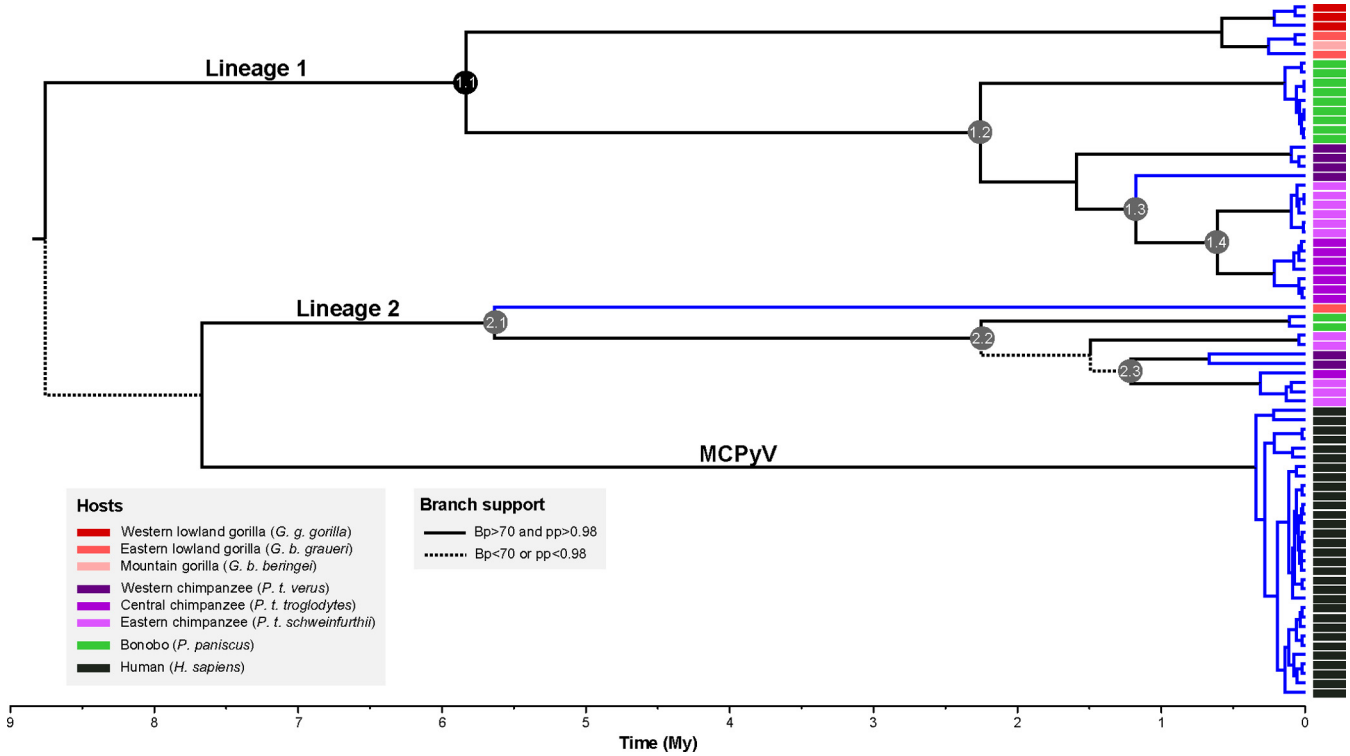

FIG 2 Chronogram derived from an alignment of partial VP1 sequences. This chronogram was obtained through BMCMC analyses run under a multispecies coalescent model (the clades corresponding to entities considered species are highlighted in blue). Other BMCMC analyses run under different tree priors and ML analyses gave similar results. The root of the tree was the most frequently observed in all posterior samples of trees (posterior probability of ca. 0.60) and was also retrieved by rooting the ML tree at its center. The six gray circles stand for the main nodes whose date estimates are given in full in Tables 5 and 6; the black circle indicates the node that was used to calibrate the analyses. Note that these circles coincide with putative codivergence events. Bp, bootstrap; pp, posterior probability.

TABLE 3 Results of BaTS tests for host specificity[a]

Host species or subspecies (no. in group)	Mean association index	Mean parsimony score	Mean maximum exclusive single-state clade size[b]	*P* value
Species (5)	0.016	6		0
G. beringei			3	<0.002
G. gorilla			3	<0.002
H. sapiens			31	<0.002
P. paniscus			9	<0.002
P. troglodytes			17	<0.002
Subspecies (8)	0.4	11		0
G. beringei beringei			1	1
G. beringei graueri			1	1
G. gorilla gorilla			3	<0.002
H. sapiens			31	<0.002
P. paniscus			9	<0.002
P. troglodytes schweinfurthii			6	<0.002
P. troglodytes troglodytes			7	<0.002
P. troglodytes verus			3	<0.002

[a] The values reported are derived from analyses performed on posterior sets of trees generated under the 14-species coalescent model. Values were very similar when posterior samples of trees obtained under a constant population size coalescent model or a birth-death speciation model were analyzed.
[b] Number of sequences.

events, irrespective of the host taxonomic level and cost set, e.g., 10 codivergence events versus 5 noncodivergence events considering host subspecies and all cost sets. Codivergence may therefore be the dominant process at play, accompanied by less frequent noncodivergence events, e.g., the viral lineage duplication event that gave rise to lineages 1 and 2.

On short timescales, host relatedness may influence viral transmission in such a way that topological congruence ensues in the absence of real codivergence, e.g., if host jumps are facilitated by host phylogenetic proximity (the preferential host switch hypothesis) (38, 52). A further step in validating codivergence events consists of showing that host and virus divergence events are synchronized. This requires branch lengths to be properly estimated

TABLE 4 Results of reconciliation analyses with Jane

Host phylogeny	Cost set	No. of events[a]		*P* value
		Cospeciation	Not cospeciation	
Species level	1	5	2	0.056
	2	5	2	0.016
	3	5	2	0.066
Subspecies level	1	10	5	0
	2	10	5	0
	3	10	5	0

[a] For the solution which was the most parsimonious in number of events.

TABLE 5 Absolute times to the most recent common ancestors of PyV in lineages 1 and 2

Statistical framework or reference	Diversification model or smoothing factor[a]	Median tMRCA (10^6 yr) or ML estimate (10^6 yr; 95% HPD or Bp interval)[b]						
		Lineage 1				Lineage 2		
		Node 1.1 (all)[c]	Node 1.2 (panine)	Node 1.3 (*P. troglodytes*)	Node 1.4 (*P. troglodytes troglodytes* + *P. troglodytes schweinfurthii*)	Node 2.1 (all)	Node 2.2 (panine)	Node 2.3 (*P. troglodytes*)
BMCMC	Coalescent, constant population size	5.62	**2.15 (1.54–2.85)**	**1.09 (0.74–1.48)**	**0.57 (0.36–0.85)**	5.36 (3.71–7.31)	**2.12 (1.47–2.92)**	**1.11 (0.75–1.55)**
	Multispecies coalescent	5.62	**2.25 (1.57–3.11)**	**1.18 (0.79–1.67)**	**0.61 (0.37–0.91)**	5.63 (3.84–7.95)	**2.25 (1.52–3.18)**	**1.21 (0.79–1.73)**
	Speciation, birth-death	5.62	**2.06 (1.46–2.72)**	**1.05 (0.71–1.44)**	**0.54 (0.34–0.79)**	5.27 (3.58–7.20)	**2.04 (1.38–2.80)**	**1.07 (0.69–1.48)**
ML	1	5.62	**2.29 (1.61–4.56)**	**1.06 (0.76–2.22)**	**0.54 (0.23–0.84)**	5.51 (3.97–24.92)	**2.27 (1.77–25.37)**	**1.21 (0.75–13.33)**
	100	5.62	**2.24 (1.61–4.19)**	**1.04 (0.70–2.27)**	**0.53 (0.18–1.93)**	5.54 (4.02–9.84)	**2.26 (1.79–4.98)**	**1.21 (0.76–2.24)**
Prado-Martinez et al. (43)[d]		5.62	0.87	0.42	0.17	5.6	0.87	0.42

[a] Diversification models were used for Bayesian Markov chain Monte Carlo (BMCMC) analyses; smoothing factors were used for maximum likelihood (ML) analyses (under the MG94-like model of codon evolution).
[b] Estimates that are incompatible with those determined in Prado-Martinez et al. (43) appear in bold. The 95% highest posterior density (HPD) was used for BMCMC analyses; bootstrap (Bp) intervals were used for ML analyses. Bootstrap intervals were determined using 100 bootstrap pseudoreplicates of the codon data set from which branch lengths were reestimated on the ML topology; all trees were rooted using TempEst. tMRCA, time to the most recent common ancestor.
[c] The respective node was used to calibrate the trees.
[d] Assuming a mutation rate of 1e−9 mutations/base pair/year.

throughout the phylogeny. Here, we speculate that the well-documented time dependency of molecular rates, which posits an apparent decay of molecular rates with increasing measurement timescales (19, 45–47), may have resulted in overestimating recent divergence dates derived from our initial molecular clock analyses which were calibrated with an ancient divergence event. In line with this hypothesis, we found that the relative timescales of host and virus divergence events were in good agreement when these estimates were rescaled using an arbitrary unit set to a recent divergence event, i.e., a procedure similar to calibrating the molecular clock with this recent divergence event. In addition, codivergence events were also synchronous in the viral lineages 1 and 2.

Overall, we observed (i) marked host specificity, (ii) frequent codivergence events, and (iii) the synchronicity of a number of codivergence events. The evolution of MCPyV-related viruses with their hominine hosts therefore appears to have been mostly driven by host-PyV codivergence. A number of other human PyV have been shown to be closely related to great ape PyV (22, 53–56). The respective lineages may represent promising opportunities to test whether the dominance of codivergence events can be generalized throughout the PyV family tree. Regardless, the findings reported here lend support to the hypothesis of an ancient association of PyV and their animal hosts, which the well-known separation of mammal and bird PyV and the recent discovery of the first fish and arthropod PyV already pinpointed (6, 8, 57). In a recently published phylogeny based on the large T antigen, the root age of the family tree was more than 11 times the age of the MRCA of MCPyV-related viruses (20). Assuming this MRCA dates back to about 6 million years ago, the family root would be more than 60 million years old. Assuming that the PyV family tree is affected by the phenomenon of time dependency of molecular rates, the root age of the family may be even more ancient, as recently suggested by C. B. Buck et al. (8).

Although a robust signal for codivergence exists, we did not observe strict codivergence of MCPyV-related viruses and their hominine hosts. For example, in our phylogenetic analyses, the placement of the MCPyV lineage is ambiguous, and the most ancient divergence event of polyomaviruses apparently postdates the respective divergence event of their hominine hosts. Although these observations may be explained by limitations of the models of sequence evolution we used, we cannot exclude the hypothesis that they reflect biological reality. Since hominine species are recently diverged, the combination of ancestral viral diversity and incomplete lineage sorting may suffice to explain apparent deviations from strict codivergence; i.e., perfect patterns of codivergence are not necessarily expected, even where no other processes have been at play (19). However, a notion emerging in the literature is that a mixture of processes, including, but not restricted to, measurable codivergence with their hosts, will generally provide a better explanation for dsDNA virus evolution in the long run than strict codivergence. For example, it was proposed that herpes simplex virus 2 (HSV-2) arose as a consequence of the transmission of a chimpanzee simplex virus to the human lineage (50). Similarly, host switches as well as lineage duplications have been documented in papillomaviruses (2, 3). It seems clear that processes other than codivergence were also at play during PyV evolution, as notably illustrated by the 13 human PyV identified thus far and

the two great ape lineages documented in this study. Further biological characterization of representatives of these lineages may reveal whether these noncodivergence events were driven by adaptive change, e.g., tissue tropism, or stochastic, e.g., demographic, processes (58).

ACKNOWLEDGMENTS

Research on Central chimpanzees, Cross River gorillas, and Western lowland gorillas in Cameroon was conducted with permission from the Ministries of Health, Forests and Wildlife and Research and benefited from the assistance of R. Ikfuingei in the field and D. Ryu in the lab. Research on Western lowland gorillas in the Central African Republic was conducted with permission from the Ministère de l'Education Nationale, de l'Alphabetisation, de l'Enseignement Supérieur et de la Recherche and benefited from the assistance of the staff of Dzanga-Sangha Protected Areas and local trackers and assistants. Research on bonobos in the Democratic Republic of the Congo (DRC) was conducted with permission from the Institut Congolais pour la Conservation de la Nature (ICCN) and benefited from the assistance of students and field assistants of the Salonga Bonobo Project. Research on Central chimpanzees and Western lowland gorillas in Gabon was conducted with permission from the Agence Nationale des Parcs Nationaux (ANPN) and the Centre de la Recherche Scientifique et Technologique and benefited from the assistance of Wildlife Conservation Society and ANPN staff as well as field assistants of the Loango Ape Project. Research on Eastern chimpanzees in Uganda was conducted with permission from the Uganda Wildlife Authority (UWA), the Uganda National Council for Science and Technology, and the Makerere University Biological Field Station (MUBFS) and benefited from the assistance of many students and assistants at the Budongo Conservation Field Station (BCFS) and MUBFS. Research on Eastern lowland and mountain gorillas in the DRC, Rwanda, and Uganda was conducted with permission from ICCN, the Rwandan Development Board, and UWA and benefited from the assistance of the Mountain Gorilla Veterinary Project, the Dian Fossey Gorilla Fund International, and Conservation Through Public Health for the organization of the censuses. We are grateful to J. Gogarten for his careful proofreading of the manuscript.

Research on Central chimpanzees, Cross River gorillas, and Western lowland gorillas in Cameroon received support from Working Dogs for Conservation, the Wildlife Conservation Society, the U.S. Fish and Wildlife Service Great Ape Conservation Fund, the Association of Zoos and Aquariums Conservation Endowment Fund, and the Agence Nationale de Recherches sur le SIDA (ANRS), France (ANRS 12125, ANRS 12182, ANRS 12555, and ANRS 12325). Research on Western lowland gorillas in the Central African Republic received support from the Primate Habituation Programme, the World Wildlife Fund (WWF), the grant agency of the Czech Republic (206/09/0927), the Institute of Vertebrate Biology of the Academy of Sciences of the Czech Republic (RVO68081766), the European Social Fund, and a Praemium Academiae award to J. Lukes. Research on Central chimpanzees and Western lowland gorillas in Gabon received support from WCS and the Société pour la Conservation et le Développement. Research on Eastern chimpanzees in Uganda received support from the Royal Zoological Society of Scotland (core funding of the BCFS). Research on Eastern lowland and mountain gorillas in the Democratic Republic of the Congo, Rwanda, and Uganda received support from WWF Sweden, the Fair Play Foundation, the Netherlands Directorate General for International Cooperation through the Greater Virunga Transboundary Collaboration, the Berggorilla & Regenwald Direkthilfe e.V., the Max Planck Society, and WCS. J.O.W. was funded by the NIH (K01-AI110181) and the University of California Laboratory Fees Research Program (grant number 12-LR-236617).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

TABLE 6 Relative times to the most recent common ancestors of PyV in lineages 1 and 2

Statistical framework or reference	Diversification model or smoothing factor[a]	Median tMRCA or ML estimate (95% HPD or Bp interval)[b]						
		Lineage 1				Lineage 2		
		Node 1.1 (all)[c]	Node 1.2 (panine)	Node 1.3 (*P. troglodytes*)	Node 1.4 (*P. troglodytes troglodytes* + *P. troglodytes schweinfurthii*)	Node 2.1 (all)	Node 2.2 (panine)	Node 2.3 (*P. troglodytes*)
BMCMC	Coalescent, constant population size	9.82	3.77 (2.70–5.00)	1.91 (1.30–2.60)	1 (0.63–1.49)	**9.40 (6.51–12.82)**	3.72 (2.58–5.12)	1.95 (1.31–2.72)
	Multispecies coalescent	9.18	3.69 (2.57–5.10)	1.93 (1.29–2.74)	1 (0.61–1.49)	**9.23 (6.29–13.03)**	3.69 (2.49–5.21)	1.98 (1.29–2.84)
	Speciation, birth-death	10.37	3.81 (2.70–5.04)	1.94 (1.31–2.67)	1 (0.63–1.46)	**9.76 (6.63–13.33)**	3.78 (2.56–5.18)	1.98 (1.28–2.74)
ML	1	10.37	4.24 (2.98–8.44)	1.96 (1.41–4.11)	1 (0.43–1.58)	10.20 (7.35–46.15)	4.20 (3.28–46.98)	2.24 (1.39–24.68)
	100	10.60	4.22 (3.04–7.90)	1.96 (1.32–4.28)	1 (0.34–3.64)	**10.45 (7.58–18.57)**	4.26 (3.38–9.40)	2.28 (1.43–4.22)
Prado-Martinez et al. (43)		32.11	4.98	2.39	1	32.11	4.98	2.39

[a] Diversification models were used for Bayesian Markov chain Monte Carlo (BMCMC) analyses; smoothing factors were used for maximum likelihood (ML) analyses (under the MG94-like model of codon evolution).

[b] Estimates that are incompatible with those determined in Prado-Martinez et al. (43) appear in bold. For time to the most recent common ancestor (tMRCA), 1 unit = tMRCA (*P. troglodytes troglodytes* + *P. troglodytes schweinfurthii*). The 95% highest posterior density (HPD) was used for BMCMC analyses; bootstrap (Bp) intervals were used for ML analyses. Bootstrap intervals were determined using 100 bootstrap pseudoreplicates of the codon data set from which branch lengths were reestimated on the ML topology; all trees were rooted using TempEst.

[c] No values for highest posterior density or bootstrap interval because this node was used to calibrate the trees.

FUNDING INFORMATION

This work, including the efforts of Klara Judita Petrzelkova, was funded by Academy of Sciences of the Czech Republic (RVO68081766). This work, including the efforts of Klara Judita Petrzelkova, was funded by Grant Agency of the Czech Republic (206/09/0927). This work, including the efforts of Joel O. Wertheim, was funded by University of California Laboratory Fees Research Program (12-LR-236617). This work, including the efforts of Joel O. Wertheim, was funded by HHS | National Institutes of Health (NIH) (K01-AI110181). This work, including the efforts of Martine Peeters, was funded by Agence Nationale de Recherches sur le Sida et les Hepatites Virales (ANRS) (12125, 12182, 12555, and 12325).

REFERENCES

1. **Compton AA, Emerman M.** 2013. Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. PLoS Pathog **9:**e1003135. http://dx.doi.org/10.1371/journal.ppat.1003135.
2. **Garcia-Perez R, Ibanez C, Godinez JM, Arechiga N, Garin I, Perez-Suarez G, de Paz O, Juste J, Echevarria JE, Bravo IG.** 2014. Novel papillomaviruses in free-ranging Iberian bats: no virus-host co-evolution, no strict host specificity, and hints for recombination. Genome Biol Evol **6:**94–104. http://dx.doi.org/10.1093/gbe/evt211.
3. **Gottschling M, Goker M, Stamatakis A, Bininda-Emonds OR, Nindl I, Bravo IG.** 2011. Quantifying the phylodynamic forces driving papillomavirus evolution. Mol Biol Evol **28:**2101–2113. http://dx.doi.org/10.1093/molbev/msr030.
4. **Lavergne A, Donato D, Gessain A, Niphuis H, Nerrienet E, Verschoor EJ, Lacoste V.** 2014. African great apes are naturally infected with roseoloviruses closely related to human herpesvirus 7. J Virol **88:**13212–13220. http://dx.doi.org/10.1128/JVI.01490-14.
5. **McGeoch DJ, Rixon FJ, Davison AJ.** 2006. Topics in herpesvirus genomics and evolution. Virus Res **117:**90–104. http://dx.doi.org/10.1016/j.virusres.2006.01.002.
6. **Johne R, Buck CB, Allander T, Atwood WJ, Garcea RL, Imperiale MJ, Major EO, Ramqvist T, Norkin LC.** 2011. Taxonomical developments in the family *Polyomaviridae*. Arch Virol **156:**1627–1634. http://dx.doi.org/10.1007/s00705-011-1008-x.
7. **Mishra N, Pereira M, Rhodes RH, An P, Pipas JM, Jain K, Kapoor A, Briese T, Faust PL, Lipkin WI.** 2014. Identification of a novel polyomavirus in a pancreatic transplant recipient with retinal blindness and vasculitic myopathy. J Infect Dis **210:**1595–1599. http://dx.doi.org/10.1093/infdis/jiu250.
8. **Buck CB, Van Doorslaer K, Peretti A, Geoghegan EM, Tisza MJ, An P, Katz JP, Pipas JM, McBride AA, Camus AC, McDermott AJ, Dill JA, Delwart E, Ng TF, Farkas K, Austin C, Kraberger S, Davison W, Pastrana DV, Varsani A.** 2016. The ancient evolutionary history of polyomaviruses. PLoS Pathog **12:**e1005574. http://dx.doi.org/10.1371/journal.ppat.1005574.
9. **Rockett RJ, Bialasiewicz S, Mhango L, Gaydon J, Holding R, Whiley DM, Lambert SB, Ware RS, Nissen MD, Grimwood K, Sloots TP.** 2015. Acquisition of human polyomaviruses in the first 18 months of life. Emerg Infect Dis **21:**365–367. http://dx.doi.org/10.3201/eid2102.141429.
10. **Feng H, Shuda M, Chang Y, Moore PS.** 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science **319:**1096–1100. http://dx.doi.org/10.1126/science.1152586.
11. **Ho J, Jedrych JJ, Feng H, Natalie AA, Grandinetti L, Mirvish E, Crespo MM, Yadav D, Fasanella KE, Proksell S, Kuan SF, Pastrana DV, Buck CB, Shuda Y, Moore PS, Chang Y.** 2015. Human polyomavirus 7-associated pruritic rash and viremia in transplant recipients. J Infect Dis **211:**1560–1565. http://dx.doi.org/10.1093/infdis/jiu524.
12. **van der Meijden E, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC.** 2010. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromised patient. PLoS Pathog **6:**e1001024. http://dx.doi.org/10.1371/journal.ppat.1001024.
13. **Perez-Losada M, Christensen RG, McClellan DA, Adams BJ, Viscidi RP, Demma JC, Crandall KA.** 2006. Comparing phylogenetic codivergence between polyomaviruses and their hosts. J Virol **80:**5663–5669. http://dx.doi.org/10.1128/JVI.00056-06.
14. **Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R.** 2009. Phylogenetics, evolution, and medical importance of polyomaviruses. Infect Genet Evol **9:**784–799. http://dx.doi.org/10.1016/j.meegid.2009.04.008.
15. **Tao Y, Shi M, Conrardy C, Kuzmin IV, Recuenco S, Agwanda B, Alvarez DA, Ellison JA, Gilbert AT, Moran D, Niezgoda M, Lindblade KA, Holmes EC, Breiman RF, Rupprecht CE, Tong S.** 2013. Discovery of diverse polyomaviruses in bats and the evolutionary history of the *Polyomaviridae*. J Gen Virol **94:**738–748. http://dx.doi.org/10.1099/vir.0.047928-0.
16. **Agostini HT, Yanagihara R, Davis V, Ryschkewitsch CF, Stoner GL.** 1997. Asian genotypes of JC virus in Native Americans and in a Pacific Island population: markers of viral evolution and human migration. Proc Natl Acad Sci U S A **94:**14542–14546. http://dx.doi.org/10.1073/pnas.94.26.14542.
17. **Shackelton LA, Rambaut A, Pybus OG, Holmes EC.** 2006. JC virus evolution and its association with human populations. J Virol **80:**9928–9933. http://dx.doi.org/10.1128/JVI.00441-06.
18. **Sugimoto C, Kitamura T, Guo J, Al-Ahdal MN, Shchelkunov SN, Otova B, Ondrejka P, Chollet JY, El-Safi S, Ettayebi M, Gresenguet G, Kocagoz T, Chaiyarasamee S, Thant KZ, Thein S, Moe K, Kobayashi N, Taguchi F, Yogo Y.** 1997. Typing of urinary JC virus DNA offers a novel means of tracing human migrations. Proc Natl Acad Sci U S A **94:**9191–9196. http://dx.doi.org/10.1073/pnas.94.17.9191.
19. **Sharp PM, Simmonds P.** 2011. Evaluating the evidence for virus/host co-evolution. Curr Opin Virol **1:**436–441. http://dx.doi.org/10.1016/j.coviro.2011.10.018.
20. **Calvignac-Spencer S, Feltkamp MC, Daugherty MD, Moens U, Ramqvist T, Johne R, Ehlers B.** 2016. A taxonomy update for the family *Polyomaviridae*. Arch Virol **161:**1739–1750. http://dx.doi.org/10.1007/s00705-016-2794-y.
21. **Leendertz FH, Scuda N, Cameron KN, Kidega T, Zuberbuhler K, Leendertz SA, Couacy-Hymann E, Boesch C, Calvignac S, Ehlers B.** 2011. African great apes are naturally infected with polyomaviruses closely related to Merkel cell polyomavirus. J Virol **85:**916–924. http://dx.doi.org/10.1128/JVI.01585-10.
22. **Scuda N, Madinda NF, Akoua-Koffi C, Adjogoua EV, Wevers D, Hofmann J, Cameron KN, Leendertz SA, Couacy-Hymann E, Robbins M, Boesch C, Jarvis MA, Moens U, Mugisha L, Calvignac-Spencer S, Leendertz FH, Ehlers B.** 2013. Novel polyomaviruses of nonhuman primates: genetic and serological predictors for the existence of multiple unknown polyomaviruses within the human population. PLoS Pathog **9:**e1003429. http://dx.doi.org/10.1371/journal.ppat.1003429.
23. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.
24. **Gouy M, Guindon S, Gascuel O.** 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol **27:**221–224. http://dx.doi.org/10.1093/molbev/msp259.
25. **Talavera G, Castresana J.** 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol **56:**564–577. http://dx.doi.org/10.1080/10635150701472164.
26. **Martin DP, Murrell B, Golden M, Khoosal A, Muhire B.** 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol **1:**vev003. http://dx.doi.org/10.1093/ve/vev003.
27. **Darriba D, Taboada GL, Doallo R, Posada D.** 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods **9:**772. http://dx.doi.org/10.1038/nmeth.2109.
28. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59:**307–321. http://dx.doi.org/10.1093/sysbio/syq010.
29. **Guindon S, Lethiec F, Duroux P, Gascuel O.** 2005. PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res **33:**W557–W559. http://dx.doi.org/10.1093/nar/gki352.
30. **Rambaut A, Lam TT, Max Carvalho L, Pybus OG.** 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol **2:**vev007.
31. **Drummond AJ, Suchard MA, Xie D, Rambaut A.** 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol **29:**1969–1973. http://dx.doi.org/10.1093/molbev/mss075.

32. **Heled J, Drummond AJ.** 2010. Bayesian inference of species trees from multilocus data. Mol Biol Evol **27:**570–580. http://dx.doi.org/10.1093/molbev/msp274.
33. **Parker J, Rambaut A, Pybus OG.** 2008. Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. Infect Genet Evol **8:**239–246. http://dx.doi.org/10.1016/j.meegid.2007.08.001.
34. **Fujisawa T, Barraclough TG.** 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. Syst Biol **62:**707–724. http://dx.doi.org/10.1093/sysbio/syt033.
35. **Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP.** 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Syst Biol **55:**595–609. http://dx.doi.org/10.1080/10635150600852011.
36. **Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R.** 2010. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms Mol Biol **5:**16. http://dx.doi.org/10.1186/1748-7188-5-16.
37. **Shimodaira H, Hasegawa M.** 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics **17:**1246–1247. http://dx.doi.org/10.1093/bioinformatics/17.12.1246.
38. **Charleston MA, Robertson DL.** 2002. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. Syst Biol **51:**528–535. http://dx.doi.org/10.1080/10635150290069940.
39. **Pond SL, Frost SD, Muse SV.** 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics **21:**676–679. http://dx.doi.org/10.1093/bioinformatics/bti079.
40. **Sanderson MJ.** 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics **19:**301–302. http://dx.doi.org/10.1093/bioinformatics/19.2.301.
41. **Muse SV, Gaut BS.** 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol Biol Evol **11:**715–724.
42. **Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL.** 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol **32:**1342–1353. http://dx.doi.org/10.1093/molbev/msv022.
43. **Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, Cagan A, Theunert C, Casals F, Laayouni H, Munch K, Hobolth A, Halager AE, Malig M, Hernandez-Rodriguez J, Hernando-Herraez I, Prufer K, Pybus M, Johnstone L, Lachmann M, Alkan C, Twigg D, Petit N, Baker C, Hormozdiari F, Fernandez-Callejo M, Dabad M, Wilson ML, Stevison L, Camprubi C, Carvalho T, Ruiz-Herrera A, Vives L, Mele M, Abello T, Kondova I, Bontrop RE, Pusey A, Lankester F, Kiyang JA, Bergl RA, Lonsdorf E, Myers S, Ventura M, Gagneux P, Comas D, et al.** 2013. Great ape genetic diversity and population history. Nature **499:**471–475. http://dx.doi.org/10.1038/nature12228.
44. **Roy J, Arandjelovic M, Bradley BJ, Guschanski K, Stephens CR, Bucknell D, Cirhuza H, Kusamba C, Kyungu JC, Smith V, Robbins MM, Vigilant L.** 2014. Recent divergences and size decreases of eastern gorilla populations. Biol Lett **10:**20140811. http://dx.doi.org/10.1098/rsbl.2014.0811.
45. **Aiewsakun P, Katzourakis A.** 1 June 2016. Time-dependent rate phenomenon in viruses. J Virol http://dx.doi.org/10.1128/jvi.00593-16.
46. **Duchene S, Holmes EC, Ho SY.** 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc Biol Sci **281:**20140732. http://dx.doi.org/10.1098/rspb.2014.0732.
47. **Ho SY, Phillips MJ, Cooper A, Drummond AJ.** 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol Biol Evol **22:**1561–1568. http://dx.doi.org/10.1093/molbev/msi145.
48. **Wertheim JO, Chu DK, Peiris JS, Kosakovsky Pond SL, Poon LL.** 2013. A case for the ancient origin of coronaviruses. J Virol **87:**7039–7045. http://dx.doi.org/10.1128/JVI.03273-12.
49. **Wertheim JO, Kosakovsky Pond SL.** 2011. Purifying selection can obscure the ancient age of viral lineages. Mol Biol Evol **28:**3355–3365. http://dx.doi.org/10.1093/molbev/msr170.
50. **Wertheim JO, Smith MD, Smith DM, Scheffler K, Kosakovsky Pond SL.** 2014. Evolutionary origins of human herpes simplex viruses 1 and 2. Mol Biol Evol **31:**2356–2364. http://dx.doi.org/10.1093/molbev/msu185.
51. **Feltkamp MC, Kazem S, van der Meijden E, Lauber C, Gorbalenya AE.** 2013. From Stockholm to Malawi: recent developments in studying human polyomaviruses. J Gen Virol **94:**482–496. http://dx.doi.org/10.1099/vir.0.048462-0.
52. **Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE.** 2010. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science **329:**676–679. http://dx.doi.org/10.1126/science.1188836.
53. **Deuzing I, Fagrouch Z, Groenewoud MJ, Niphuis H, Kondova I, Bogers W, Verschoor EJ.** 2010. Detection and characterization of two chimpanzee polyomavirus genotypes from different subspecies. Virol J **7:**347. http://dx.doi.org/10.1186/1743-422X-7-347.
54. **Johne R, Enderlein D, Nieper H, Muller H.** 2005. Novel polyomavirus detected in the feces of a chimpanzee by nested broad-spectrum PCR. J Virol **79:**3883–3887. http://dx.doi.org/10.1128/JVI.79.6.3883-3887.2005.
55. **Madinda NF, Robbins MM, Boesch C, Leendertz FH, Ehlers B, Calvignac-Spencer S.** 2015. Genome sequence of a Central chimpanzee-associated polyomavirus related to BK and JC polyomaviruses, *Pan troglodytes troglodytes* polyomavirus 1. Genome Announc **3:**e00888–15. http://dx.doi.org/10.1128/genomeA.00888-15.
56. **van Persie J, Buitendijk H, Fagrouch Z, Bogers W, Haaksma T, Kondova I, Verschoor EJ.** 2016. Complete genome sequence of a novel chimpanzee polyomavirus from a western common chimpanzee. Genome Announc **4:**e01406–15. http://dx.doi.org/10.1128/genomeA.01406-15.
57. **Peretti A, FitzGerald PC, Bliskovsky V, Pastrana DV, Buck CB.** 2015. Genome sequence of a fish-associated polyomavirus, black sea bass (Centropristis striata) polyomavirus 1. Genome Announc **3:**e01476–14. http://dx.doi.org/10.1128/genomeA.01476-14.
58. **Anthony SJ, Islam A, Johnson C, Navarrete-Macias I, Liang E, Jain K, Hitchens PL, Che X, Soloyvov A, Hicks AL, Ojeda-Flores R, Zambrana-Torrelio C, Ulrich W, Rostal MK, Petrosov A, Garcia J, Haider N, Wolfe N, Goldstein T, Morse SS, Rahman M, Epstein JH, Mazet JK, Daszak P, Lipkin WI.** 2015. Non-random patterns in viral diversity. Nat Commun **6:**8147. http://dx.doi.org/10.1038/ncomms9147.