



# Discovery and Characterization of Distinct Simian Pegiviruses in Three Wild African Old World Monkey Species

Samuel D. Sibley<sup>1</sup>, Michael Lauck<sup>2</sup>, Adam L. Bailey<sup>2</sup>, David Hyeroba<sup>3</sup>, Alex Tumukunde<sup>3</sup>, Geoffrey Weny<sup>3</sup>, Colin A. Chapman<sup>3,4</sup>, David H. O'Connor<sup>2,5</sup>, Tony L. Goldberg<sup>1,3,5</sup>, Thomas C. Friedrich<sup>1,5\*</sup>

**1** Department of Pathobiological Sciences, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **2** Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, Wisconsin, United States of America, **3** Makerere University, Kampala, Uganda, **4** Department of Anthropology and McGill School of Environment, Montreal, Quebec, Canada, **5** Wisconsin National Primate Research Center, Madison, Wisconsin, United States of America

## Abstract

Within the *Flaviviridae*, the recently designated genus *Pegivirus* has expanded greatly due to new discoveries in bats, horses, and rodents. Here we report the discovery and characterization of three simian pegiviruses (SPgV) that resemble human pegivirus (HPgV) and infect red colobus monkeys (*Procolobus tephrosceles*), red-tailed guenons (*Cercopithecus ascanius*) and an olive baboon (*Papio anubis*). We have designated these viruses SPgV<sub>krc</sub>, SPgV<sub>krtg</sub> and SPgV<sub>kbab</sub>, reflecting their host species' common names, which include reference to their location of origin in Kibale National Park, Uganda. SPgV<sub>krc</sub> and SPgV<sub>krtg</sub> were detected in 47% (28/60) of red colobus and 42% (5/12) red-tailed guenons, respectively, while SPgV<sub>kbab</sub> infection was observed in 1 of 23 olive baboons tested. Infections were not associated with any apparent disease, despite the generally high viral loads observed for each variant. These viruses were monophyletic and equally divergent from HPgV and pegiviruses previously identified in chimpanzees (SPgV<sub>cpz</sub>). Overall, the high degree of conservation of genetic features among the novel SPgVs, HPgV and SPgV<sub>cpz</sub> suggests conservation of function among these closely related viruses. Our study describes the first primate pegiviruses detected in Old World monkeys, expanding the known genetic diversity and host range of pegiviruses and providing insight into the natural history of this genus.

**Citation:** Sibley SD, Lauck M, Bailey AL, Hyeroba D, Tumukunde A, et al. (2014) Discovery and Characterization of Distinct Simian Pegiviruses in Three Wild African Old World Monkey Species. PLoS ONE 9(6): e98569. doi:10.1371/journal.pone.0098569

**Editor:** Ilhem Messaoudi, University of California Riverside, United States of America

**Received:** October 2, 2014; **Accepted:** May 5, 2014; **Published:** June 11, 2014

**Copyright:** © 2014 Sibley et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was funded by National Institutes of Health (NIH) grants TW009237, as part of the joint NIH-NSF Ecology of Infectious Disease program, and R01 AI084787; the UK Economic and Social Research Council, through grants P51OD011106 and P51RR000167 to the Wisconsin National Primate Research Center; and by the University of Wisconsin School of Medicine and Public Health Wisconsin Partnership Program through the Wisconsin Center for Infectious Disease (WisCID). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: thomasf@primate.wisc.edu

## Introduction

Pegiviruses are single-stranded, positive-sense RNA viruses within the family *Flaviviridae*. This family includes three additional genera, *Hepacivirus*, *Pestivirus*, and *Flavivirus*, and encompasses a diverse set of viruses, including important human pathogens such as hepatitis C virus (HCV), a hepacivirus first described in 1989 [1]. The discovery in 1995 of human pegivirus (HPgV) [2], formerly known as GB virus C or hepatitis G virus [3,4], was the product of efforts to uncover hepatitis viruses related to HCV. While HPgV is the human virus most closely related to HCV, it does not cause hepatitis [5]. Other early work to discover HCV-like viruses focused on nonhuman primates and revealed pegiviruses distantly related to HPgV in sera from six New World monkey species [6,7] and variants closely related to HPgV in sera from captive and wild chimpanzees [2,8,9]. After 1998, few new pegiviruses were reported until 2013, a year that witnessed a flourishing of pegivirus discoveries in horses [10,11], bats [12,13], and rodents [14,15]. These recent revelations suggest that pegivirus infection may be widespread among mammals, although, until now, no related viruses had been identified in Old World monkeys.

Among the pegiviruses, HPgV has garnered the greatest attention. HPgV establishes an asymptomatic infection and circulates as at least five closely related, yet phylogenetically distinct, genotypes that cluster geographically [16]. HPgV is thought to be predominately lymphotropic, and the virus is produced *in vitro* by T- and B-lymphocytes obtained from infected individuals (reviewed in [3]). HPgV infection may persist for decades, although the majority of infections are cleared within two years in healthy individuals [17]. Unlike HCV infection, HPgV viremia is not usually coincident with detectable antibodies against envelope glycoprotein protein 2 (E2), suggesting that HPgV infection is resolved efficiently once a humoral immune response to E2 is mounted [18]. Based on detection of the virus in blood donors, the prevalence of HPgV infection is significantly lower in developed (1–5%) versus developing (~20%) countries [19]. Higher HPgV prevalence (~20–40%) is associated with sexual and parenteral risk behaviors and, accordingly, with human immunodeficiency virus type 1 (HIV) and HCV infections [20–23]. Although the cellular tropism of pegiviruses infecting animals is largely undetermined [3], the characteristics of pegivirus infection in nonhuman hosts appear to closely resemble those

observed in humans: virus can be detected at high titer in blood, infection may persist for some time, and a remarkably low degree of within-host viral genetic diversity is observed in time-series samples [7,19,24]. Interestingly, Theiler's disease, an infectious hepatitis in horses, is the only disease associated with a pegivirus infection [10].

In the past decade, research on HPgV has developed from understanding the virus's evolution and phylogeography [24–26] to clarifying its medical importance in relation to HIV infection. In HIV-positive individuals, HPgV viremia has been associated with prolonged survival and a milder HIV disease course, including higher CD4+ T cell counts, lower HIV viral load, and delayed progression to AIDS (reviewed in [18]). Impacts of HPgV infection that potentially contribute to this antagonism include direct antiviral effects; altered expression of cytokines, chemokines and HIV entry receptors; and modulation of host cell signaling pathways [18,19,27]. However, the mechanisms underlying these phenomena are still not fully understood, and no tractable animal models exist to study them.

Here we report the discovery and characterization of three pegiviruses infecting red colobus monkeys (*Procolobus tephrosceles*), red-tailed guenons (*Cercopithecus ascanius*) and an olive baboon (*Papio anubis*) from Kibale National Park, Uganda. After simian pegiviruses (SPgV) identified in chimpanzees (SPgV<sub>cpz</sub>), these viruses are the closest known relatives to HPgV. Their discovery in Old World monkeys expands the known host range of the pegiviruses and provides new insights on their natural history and evolution.

## Results

### Discovery of Novel Simian Pegiviruses

This study was conducted as part of a long-term investigation of health and conservation focused on the region of Kibale National park, western Uganda [28]. Kibale is noted for its high diversity and density of non-human primates, which host a diversity of pathogens [29–34], and which interact extensively and often antagonistically with local people and domestic animals [28]. Deep sequencing of plasma RNA from wild red colobus monkeys (*Procolobus tephrosceles*), red-tailed guenons (*Cercopithecus ascanius*) and olive baboons (*Papio anubis*) revealed the presence of three distinct, host-associated viruses with greatest similarity (via BLAST) to human and chimpanzee pegiviruses. *De novo* assembly and iterative mapping of sequencing reads recovered near-complete genomes comprising a single continuous open reading frame (ORF) and partial 5'- and 3'-untranslated regions (UTR) from most positive animals (GenBank sequence accession no. KF234499 to KF234530). According to recently adopted nomenclature [3], we tentatively designated the new viruses SPgV<sub>krc</sub>, SPgV<sub>krtg</sub> and SPgV<sub>kbab</sub> to indicate their host species of origin (Kibale red colobus, Kibale red-tailed guenon and Kibale olive baboons, respectively) and their shared common ancestry with members of the *Pegivirus* genus. Sequencing statistics, detection frequencies and viral titers in plasma, measured by TaqMan qRT-PCR, are summarized in Table 1. SPgV<sub>krc</sub> and SPgV<sub>krtg</sub> were detected in 47% (28/60) of red colobus and 42% (5/12) red-tailed guenons, respectively, while SPgV<sub>kbab</sub> infection was detected in just one baboon (of 23). High viral titers were documented, consistent with observations made for PpV infections in other mammals (Fig. S1) [10,12,35]. No animals appeared clinically ill at the time of sampling, nor have any been subsequently observed with overt clinical signs. Attempts to isolate these viruses are ongoing.

### Genome Characterization of SPgV<sub>krc</sub>, SPgV<sub>krtg</sub> and SPgV<sub>kbab</sub>

Like all *Flaviviridae*, the genomes of the newly discovered Kibale SPgVs encode a single, continuous ORF (8552 to 8697 nucleotides, nt). Cleavage site prediction along the polyprotein distinguished nine mature proteins representing the typical proposed *Pegivirus* genome organization: E1/E2/p7/NS2/NS3/NS4A/NS4B/NS5A/NS5B (see below). The translation initiation codon is poorly defined for several pegiviruses [3,9,13]. For each of the Kibale SPgVs, the first Met in-frame with the coding sequence aligned well with the experimentally determined translation initiation codon for HPgV [36]. Measured from this position, the aligned coding sequence lengths for the Kibale SPgVs were similar to those reported for SPgV<sub>cpz</sub> (8484 nt) and HPgV (8529 to 8565 nt) but were considerably shorter than those of the New World simian pegiviruses (8865–9018 nt) and the divergent pegiviruses recently reported in bats (BPgV; 8919–9777 nt), rodents (RPgV; 9855 nt) and horses (EPgV; 9570–9918).

Partial 5'UTR sequences of 498, 513 and 521 nt were recovered by deep sequencing for SPgV<sub>krtg</sub>, SPgV<sub>kbab</sub>, and SPgV<sub>krc</sub>, respectively. These sequences demonstrated significant identity to the 554-nt HPgV 5'UTR. Accordingly, structures analogous to stem-loops II, IIIa, IIIb, IVa, IVb, Va and Vb in the HPgV 5'UTR internal ribosome entry site [36,37] could be confidently resolved via RNAfold and RNAalifold analyses [38,39] (Fig. S2). Partial 3'UTR sequences of 147, 162 and 308 nt were recovered for SPgV<sub>krtg</sub>, SPgV<sub>kbab</sub>, and SPgV<sub>krc</sub>, respectively. Compared with the 312-nt HPgV 3'UTR, conservation between the Kibale SPgVs, HPgV and SPgV<sub>cpz</sub> was strong within the first ~110 nt. In particular, a pyrimidine-rich poly(C) region, reminiscent of motifs potentially involved with RNA replication and/or stabilization observed in the 3'UTRs of HCV [40–42] and of several mammalian mRNAs [43], was well conserved among the Old World primate pegiviruses. Surrounding the poly(C) region, 3'UTR structures analogous to proposed structural motifs V through VII (as defined in [44]) were apparent, with the poly(C) region always beginning in the loop of structure VI (Fig. S2). Attempts to amplify full 5'- and 3'UTR sequences via RACE and 5'-3' ligation [45] were unsuccessful.

Mature proteins encoded by the 5' half of *Flaviviridae* genomes are cleaved from the polyprotein by host signal peptidase and were assigned for the Kibale SPgVs through manual [46] and *in silico* [47,48] signalase cleavage site prediction. Proposed signal-sequence cleavage sites were well conserved among the Kibale SPgVs, HPgV and SPgV<sub>cpz</sub> (Fig. 1A). Accordingly, two envelope proteins (E1 and E2) and a p7 protein (~7.5 kDa; ExpPASy server [49]) were predicted for these new SPgVs in this region. As observed for other pegiviruses, the Kibale SPgVs apparently lack sequence encoding a core (i.e., nucleocapsid) protein [3,9], and no alternate reading frame proteins were detected. No major insertions or deletions were observed in structural proteins for the new SPgVs relative to HPgV and SPgV<sub>cpz</sub>. This differs from New World SPgVs, BPgVs, RPgVs and EPgVs, viruses with large insertions (107–183 amino acids, aa) near the C-terminus of E2 and for which signalase cleavage sites in this region are unresolved [13,14,50]. It has been speculated that the inserts in these viruses may, in fact, constitute an additional structural protein, designated "X" [13,14].

Reduced glycosylation of envelope proteins is a notable feature distinguishing HPgV from the hepaciviruses HCV and GB virus B, whose envelopes are heavily glycosylated [3,19]. Within E1, a single potential glycosylation site (N<sub>135</sub> with reference to, *wrt*, the SPgV<sub>krc</sub> ORF) was predicted for the Kibale SPgVs. This site was conserved in roughly the same location among all identified

**Table 1.** Detection frequency, TaqMan qRT-PCR titer and genome statistics for Kibale SPgVs.

	ORF (nt)	Detection Frequency	Sequenced Reads <sup>a</sup>	Ave. Coverage Depth <sup>a</sup>	Sequence Length (nt) <sup>a</sup>	Titer (g.c. mL <sup>-1</sup> ) <sup>a</sup>
SPgV <sub>kbab</sub>	8652	1/23	29023	402	9434	2.4 × 10 <sup>7</sup>
SPgV <sub>krc</sub>	8697	28/60	548–312763	5–6264	8776–9596	1.8 × 10 <sup>4</sup> –2.2 × 10 <sup>8</sup>
SPgV <sub>krtg</sub>	8631	5/12	1321–6630	21–101	9119–9284	2.1 × 10 <sup>6</sup> –2.1 × 10 <sup>7</sup>

<sup>a</sup>Ranges for positive animals; assays were sensitive to between 10 and 100 genome copies (g.c.) per 20-μL qRT-PCR reaction.  
doi:10.1371/journal.pone.0098569.t001

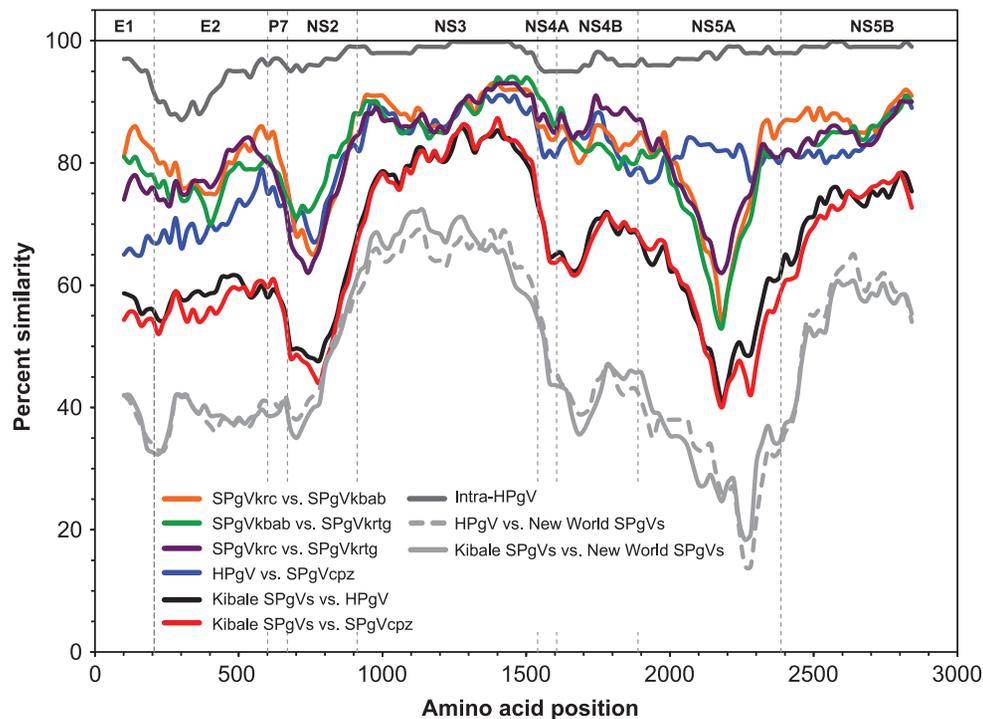
pegiviruses, and no pegivirus examined had more than two predicted potential E1 glycosylation sites. Across E2, a single potential glycosylation site was conserved among nearly all pegiviruses (N<sub>241</sub> *wrt* SPgV<sub>krc</sub>), and a second site was conserved

among all Old World primate pegiviruses (N<sub>400</sub> *wrt* SPgV<sub>krc</sub>). In total, the Kibale SPgVs each exhibited three potential E2 glycosylation sites, similar to HPgV (three to four), SPgV<sub>cpz</sub> (three) and New World SPgVs (three to four). Excluding the putative “X”

**A**

Virus	Start/E1*	E1/E2	E2/P7	P7/NS2	NS2/NS3	NS3/NS4A	NS4A/NS4B	NS4B/NS5A	NS5A/NS5B
HPgV	PATHA CGANG	GMSQG APASV	PAVEA AVAGE	SATRG RTSVL	GFVPT APVVI	EGYVRC DAGP	RPGGES APSD	DEVQVG YVWD	ROETLA SFSY
SPgV <sub>cpz</sub>	..S.. .RVGS	..AAS R.TGD	L.ALG ..V..	.....	.....	.....	..AH.	..A.. ..I..	..C...T .....
SPgV <sub>kbab</sub>	SGAGA LLSPA	..AD. ....	ISA.. .LHHG	GFS.. .S...	.....	..FA..	....	P..A.A ..N. S.TT..	..A.. PK..SC .Y..
SPgV <sub>krtg</sub>	SGVGA LLSPA	..AR. ..VAD	STA.. .LHHC	GLG..	.....	..FA..	....	P..A.A ..N. S.TH..	..A.. PK..AC .Y..
SPgV <sub>krc</sub>	QVAGG LLSPA	..AS. ..V..	ITA.. .LHHG	G....	.....	..FQ..	....	H..S.T ..N. S.TT..	.... .K..AC .Y..

▲ 15/16      ▲ 205/206      ▲ 579/580      ▲ 647/648      ▲ 885/886      ▲ 1512/1513      ▲ 1573/1574      ▲ 1853/1854      ▲ 2334/2335  
**Predicted cleavage sites**

**B**

**Figure 1. Polypeptide cleavage sites and amino acid similarity of the Kibale SPgVs and related primate pegiviruses.** (A) Amino acid sequences of the Kibale SPgVs and related viruses adjacent to predicted protease cleavage sites. Proposed cleavage sites for signalase (black triangles), NS2–NS3 protease (gray triangle), and NS3–4A protease (white triangles) are indicated. Amino acid positions of cleavage sites in relation to SPgV<sub>krc</sub> are included below the triangles. \*The signalase cleavage site predicted for the Kibale SPgVs between Start and E1 (amino acid position 15/16) is located 5' of those predicted for HPgV and SPgV<sub>cpz</sub> (amino acid positions 22/23 and 21/22, respectively) in the sequence alignment; all other cleavage sites are aligned among these viruses. (B) Sliding window amino acid similarity among primate *Pegiviruses* across aligned coding regions (window, 200 aa; step, 20 aa). Dashed vertical lines indicate start positions of inferred viral proteins. The dark gray trace demonstrating the high similarity among known HPgV variants was included for reference.  
doi:10.1371/journal.pone.0098569.g001

protein region from this analysis, the remaining pegiviruses show similar predicted potential E2 glycosylation (one to four sites) with the exception of the phylogenetically basal clade comprising BPgVs and RPgV (see below), which exhibit six to eight predicted sites.

Non-structural proteins for the members of the family *Flaviviridae* are cleaved by viral nonstructural proteins 2 (NS2) and 3 (NS3), along with NS3's cofactor NS4A [51,52], and were identified based on alignment with identified cleavage sites for other pegiviruses [53,54] and hepaciviruses [51]. The carboxy termini of NS3 and NS4A were approximated due to the current absence of experimental data for related pegiviruses [55]. Amino acids necessary for NS2-mediated autocatalytic cleavage of NS2/NS3 (His<sub>809</sub>, Glu<sub>829</sub> and Cys<sub>850</sub>, *wrt* SPgV<sub>krc</sub>) and NS3-4A-mediated cleavage of the remaining non-structural proteins (His<sub>941</sub>, Asp<sub>965</sub> and Ser<sub>1023</sub>, *wrt* SPgV<sub>krc</sub>) were conserved among the Kibale SPgVs. In addition, residues within the NS3 helicase [56] and NS5B RNA-dependent RNA polymerase (RdRp) motifs [57] were highly conserved (91% and 90% amino acid identity, respectively) among the Kibale SPgVs.

Within NS5A, the zinc-binding motif, Cx<sup>17</sup>CxCx<sup>22</sup>C, which is conserved in nearly the same form among hepaciviruses and other related viruses and plays a role RNA replication [58], was conserved among all the pegiviruses examined. The NS5A protein of HCV is known to have intrinsically disordered regions (IDRs) that are important for many of its inferred functions, including modulating host regulatory and signaling processes [59,60]; the recently discovered guereza hepacivirus (GHV) has a similar IDR [61]. IDRs lack a well-defined three-dimensional structure under native conditions [62] but may undergo substrate-induced disorder-to-order transitions allowing for their interactions with multiple binding partners [63]. To determine if these features were conserved across the hepaciviruses and pegiviruses genera, we searched for IDRs within the sequences of all Old World primate pegiviruses. IDRs were identified in the C-terminal half of NS5A for the Kibale SPgVs (residues ~2050 to ~2210). In this ~160-aa region, 44% (SPgV<sub>krig</sub>), 67% (SPgV<sub>krc</sub>) and 78% (SPgV<sub>kbab</sub>) of the amino acids were predicted to be intrinsically disordered [62]. The same analysis also predicted disorder in the 3' end of NS5A for HPgV (73% of residues 2065 to 2175, *wrt* NC\_001710) and SPgV<sub>cpz</sub> (82% of residues 2053 to 2154; *wrt* AF070476). Using ANCHOR [64], IDRs with tendency to undergo substrate-induced disorder-to-order transitions were detected in each of the viruses examined, suggesting that the pegivirus IDRs may have the capacity to bind transiently to a range of substrates. In particular, major TNFR-associated factor 2 (TRAF2) binding motifs were detected within the IDRs of each of the Old World primate pegiviruses ( $p = 0.0043$ , [65]). The consistent identification of TRAF2 binding motifs in IDRs suggests a potential role for NS5A in modulating signal transduction from members of the tumor necrosis factor (TNF) receptor family [66,67].

### Similarity Analysis

Pairwise comparison of complete open reading frame (ORF) sequences demonstrated that the three Kibale SPgVs shared ~73% nucleotide (nt) and ~81% amino acid (aa) identity (ID) with each other but were approximately equally divergent from HPgV (63% nt ID; 65% aa ID), SPgV<sub>cpz</sub> (62% nt ID; 65% aa ID), New World SPgVs (53% nt ID; 48% aa ID), and more-divergent pegiviruses infecting bats, horses and rodents (45–54% nt ID; 34–49% aa ID). ORF sequences for Kibale SPgVs infecting the same monkey species were highly similar ( $\geq 96\%$  nt ID).

Sliding-window similarity analysis across complete ORF sequences illustrated the degree of amino acid identity of the

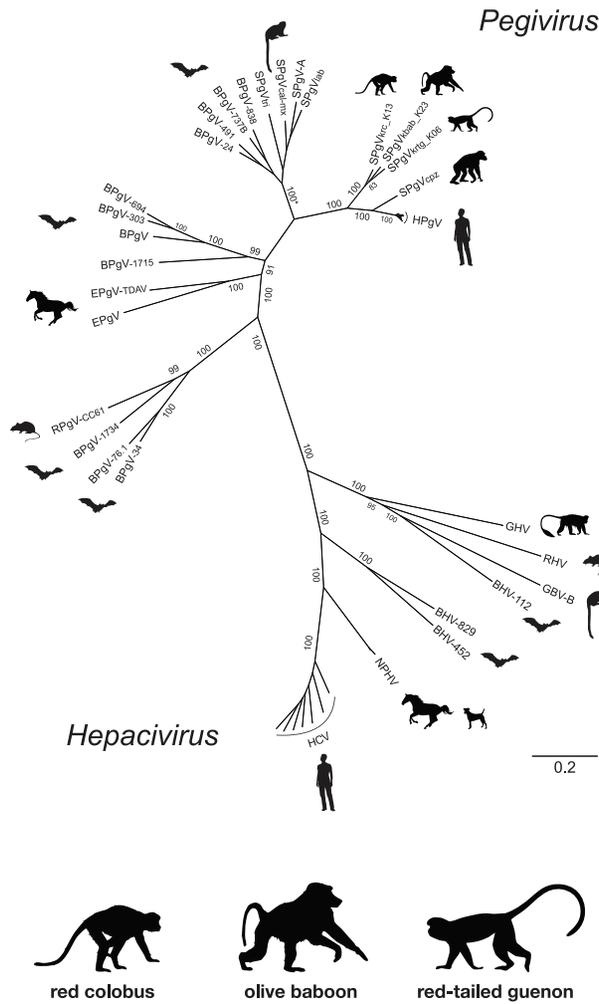
Kibale SPgVs to closely related pegiviruses (Fig. 1B). The Kibale SPgVs were essentially equally divergent from one another (Fig. 1B, orange, green and purple traces), averaging 78% ID across E1 and E2 and 89% ID within NS3. Interestingly, no Kibale SPgV pair was consistently most similar across the ORF. Together, these viruses shared greatest sequence identity with HPgV and SPgV<sub>cpz</sub> variants (Fig. 1B, black and red traces, respectively), averaging 57% ID across E1 and E2 and 81% ID in NS3. Nearly equal divergence was observed between groups comprising the New World SPgVs and the Kibale SPgVs (Fig. 1B, solid light gray trace) or the New World SPgVs and the pegiviruses detected in humans and chimpanzees (HPgV+SPgV<sub>cpz</sub> group) (Fig. 1B, dashed light gray trace). Notably, this comparison excludes variability in the C-terminus of E2, where unique insertions are present in New World SPgVs (and other divergent pegiviruses); identity of these viruses with Old World primate pegiviruses is therefore undefined in this genomic region. The greatest similarity among all the pegiviruses compared was observed over conserved NS3 helicase and NS5B RdRp motifs (17, 18), corresponding to the essential roles the encoded proteins play in the *Flaviviridae* replication cycle. Minima in sequence identity were observed in NS2 and, in particular, in the C-terminal half of NS5A, where considerable variability in sequence length and identity exists among many of the pegiviruses. This region is also where an unusual twelve-amino-acid insert was noted for the so-called "indel type" HPgVs [68]. Interestingly, the characteristic decline in NS5A sequence identity observed here and elsewhere between different pegiviruses [14] was absent between HPgV and SPgV<sub>cpz</sub> (Fig. 1B, blue trace), perhaps reflecting the close relationship between humans and chimpanzees.

### Phylogenetic Analyses

To estimate evolutionary relationships of the novel Kibale SPgVs to related viruses, we constructed a Bayesian phylogenetic tree comprising 44 ORF sequences, representing the full available genetic diversity within each major hepaciviruses and pegivirus clade (Fig. 2). This phylogeny yielded topologies consistent with established relationships among the hepaciviruses and pegiviruses [3] and demonstrates the relationships among several recently described viruses infecting horses, bats and rodents [10–12,14]. The new Kibale SPgVs are monophyletic and share a most recent common ancestor with human and chimpanzee pegiviruses. Interestingly, the evolutionary distance between the Kibale SPgVs and HPgV is considerably less than the distance observed between HCV and its closest known relative, nonprimate hepacivirus (NPHV), whose natural host is horses [69,70].

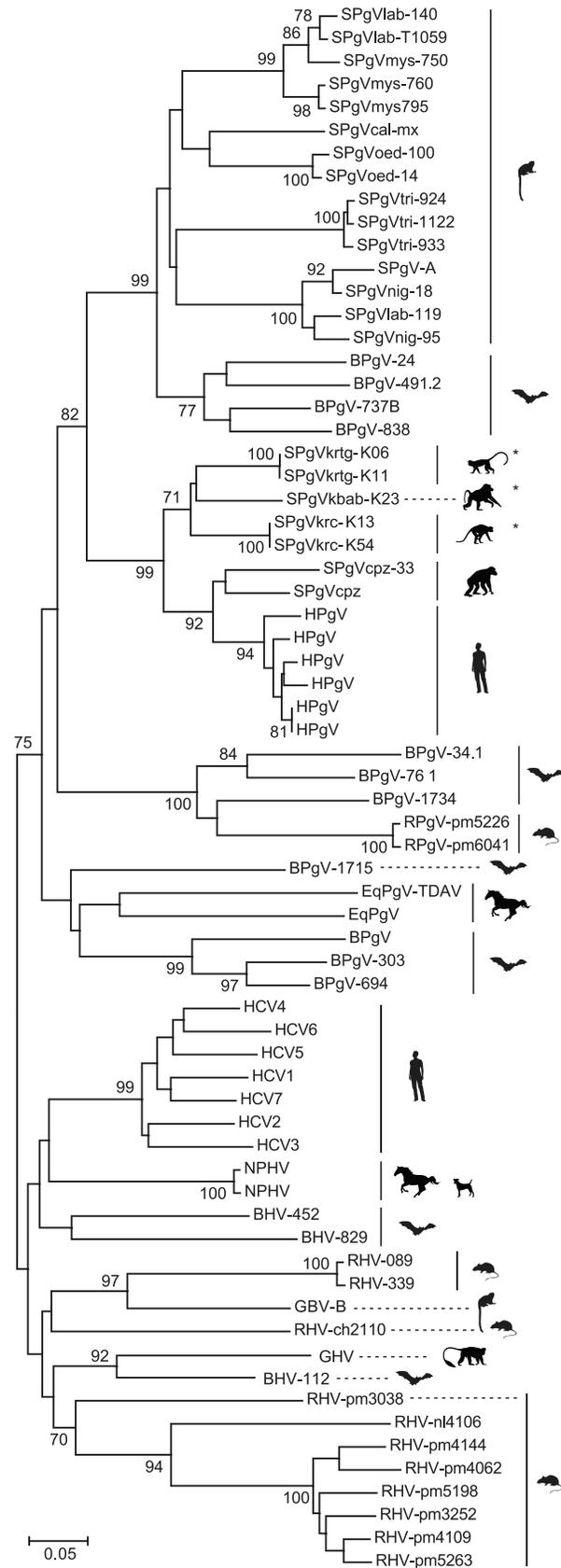
Based on our Bayesian analysis, the phylogenetic relationship among the Old World primate pegiviruses is consistent with that of their primate hosts [71], suggesting the possibility of virus-host co-speciation. However, we note that the relationship among the Kibale SPgVs varied when we used nucleotide versus amino acid alignments for phylogenetic reconstruction (not shown), an observation reflected by the relatively low posterior support obtained for the internal node of the Kibale SPgV clade (83%; Fig. 2). The relationship among the three viruses observed in the similarity plot (Fig. 1B) supports an effective polytomy among the Kibale SPgVs, in that no pair of the three Kibale SPgVs was consistently most similar across the polyprotein.

A comprehensive neighbor-joining phylogeny based on a highly conserved, 97-aa region of the NS3 helicase gene (Fig. 3), for which abundant sequence information is available for primate pegiviruses and phylogenetic comparisons are often reported, yielded a similar topology to the Bayesian tree [7–9,72,73]. Sequences included in this phylogeny encompass the full genetic



**Figure 2. Bayesian phylogeny of the Kibale simian pegiviruses, SPgVkrccr, SPgVkrtrg and SPgVkrbabr, and representative members of the Pegivirus and Hepacivirus genera.** Codon-guided sequence alignments of open reading frame protein sequences were generated via MAFFT, cleaned using Gblocks, and stripped of the third codon position, resulting in a 42-taxa, 3668-character alignment. Viruses representing the known diversity within each clade were chosen for inclusion in the Bayesian analysis. Posterior clade probabilities are shown for major branches. Scale bar indicates nucleotide substitutions per site. \*Posterior probabilities for all branches in this clade are 100. GenBank accession numbers for the included taxa are provided in Table S1  
doi:10.1371/journal.pone.0098569.g002

diversity of identified pegiviruses within each clade, with the exception of pegiviruses infecting common marmosets (*Callithrix jacchus*) [7] and several recently identified viruses infecting bats [12], for which NS3 sequences were unavailable. As documented previously [8], chimpanzee pegivirus sequences were monophyletic and considerably more divergent than the most divergent HPgVs. Additionally, these SPgVcpz variants alone exhibited diversity rivaling that observed among the three Kibale SPgVs. Overall, with the addition of the Kibale SPgVs, the between-host genetic diversity of Old World primate pegiviruses now approaches that observed for pegiviruses infecting New World primates. A striking feature of both phylogenies is the paraphyletic (i.e., multi-lineage) distribution of pegiviruses and hepaciviruses infecting bats and rodents, which confirms that these hosts are significant reservoirs for hepaciviruses and pegivirus genetic diversity [74].



**Figure 3. Neighbor-joining amino acid phylogeny of the NS3 helicase comprising 43 pegivirus and 25 hepacivirus sequences.** This 97-aa segment (polyprotein positions 1221 to 1317 relative to

HPgV, NC\_001710) is highly conserved among the *Flaviviridae* and has been targeted extensively for virus discovery and phylogenetic characterization. The sequences included in this analysis encompass the full genetic diversity of identified pegiviruses within each clade, minus those for which NS3 sequence data were unavailable, namely pegiviruses infecting the common marmoset, *Callithrix jacchus* [7] and several recently identified viruses infecting bats [12]. Inclusion of the two most diverse variants of both SPgV<sub>krc</sub> and SPgV<sub>krtg</sub> demonstrated the relatively high within-host similarity of these viruses within the study population. GenBank accession numbers for the included taxa are provided in Tables S1 and S2.  
doi:10.1371/journal.pone.0098569.g003

### Within-host Genetic Diversity

We characterized the within-host genetic diversity of each simian pegivirus detected in Kibale animals by quantifying and mapping the distributions of single nucleotide polymorphisms (SNPs) detected in deep sequencing reads along the genomic ORF. Among all variants, within-host genetic variability was low and accounted for at most 100 SNPs using a 5% SNP frequency cutoff. This level of diversity is similar to observations made for HPgV [75] and guereza hepatitis virus [61], which infects black-and-white colobus monkeys in Kibale, but is considerably lower than SNP variation observed for HCV [76].

Among the SPgVs identified in red colobus, 13 of 28 samples met our variant-calling criteria for coverage depth ( $\geq 100$  sequences) across most of the ORF ( $\geq 95\%$ ), exhibiting average coverage depths of 577 to 6264 reads. For these samples, we expect the distribution of variants detected to approximate the true within-host genetic diversity since the entire ORF was eligible for variant calling; for this reason we focused our analysis of within-host genetic diversity on this subset of SPgV<sub>krc</sub> sequences. Synonymous and nonsynonymous substitutions in this subset of samples varied from 0 to 91 (median = 15) and 0 to 9 (median = 3), respectively. To determine whether the spatial distribution of synonymous and nonsynonymous substitutions along the polyprotein was homogeneous for SPgV<sub>krc</sub>, we aggregated SNPs from these 13 samples, mapped them along the ORF and quantified the percent of total SNPs falling within each gene (Fig. 4A; Table S3). Overall, we observed a total of 318 synonymous and 55 nonsynonymous substitutions in this cohort. The distribution of synonymous substitutions was proportional to the gene length ( $R^2 = 0.97$ ; Fig. S3), and approximately 50% occurred at frequencies less than 10%. This pattern suggests a steady and essentially homogeneous contribution of random mutations to within-host nonsynonymous genetic diversity. Conversely, the distribution of nonsynonymous substitutions showed poor correlation to the gene length across the ORF ( $R^2 = 0.15$ ; Fig. S4). A clear bias existed towards nonsynonymous substitution within E2, P7, NS2, NS5A and NS5B, and nonsynonymous substitutions were absent in NS3, NS4A and NS4B.

Coverage depths for the five SPgV<sub>krtg</sub> samples (averaging 21 to 101 reads) were lower and non-uniform compared with the SPgV<sub>krc</sub> samples described above, precluding an assessment of diversity across each ORF. On the other hand, the single SPgV<sub>khab</sub> detected had an average coverage of 402 reads, with 98% ORF coverage of greater than 100 reads, yet this sample exhibited only two SNPs applying the same 5% SNP frequency cutoff, both in NS5B.

### Between-host Patterns of Selection

To assess selection along the polyprotein, we aligned three representative Kibale SPgV consensus sequences (accession no. KF234523, KF234526 and KF234530) and determined average substitution rates at non-synonymous (dN) and synonymous (dS)

sites, and their ratio (dN/dS), along a sliding window (100 aa window, 20 aa step; Fig. 4B). Consistent with observed within-host variability for SPgV<sub>krc</sub>, dS among Kibale SPgV consensus sequences was steady across the ORF, while dN showed greater variability, increasing notably within E2, P7, NS2 and NS5A and remaining steadily low throughout NS3. Accordingly, dN/dS ratios were low within NS3, suggesting that on average purifying selection acts to remove deleterious mutations from this coding region. In contrast, dN/dS ratios were  $>1$  within E2, NS2 and NS5A, suggesting that positive selection may act to favor sequence diversification in these genes among the Kibale SPgVs.

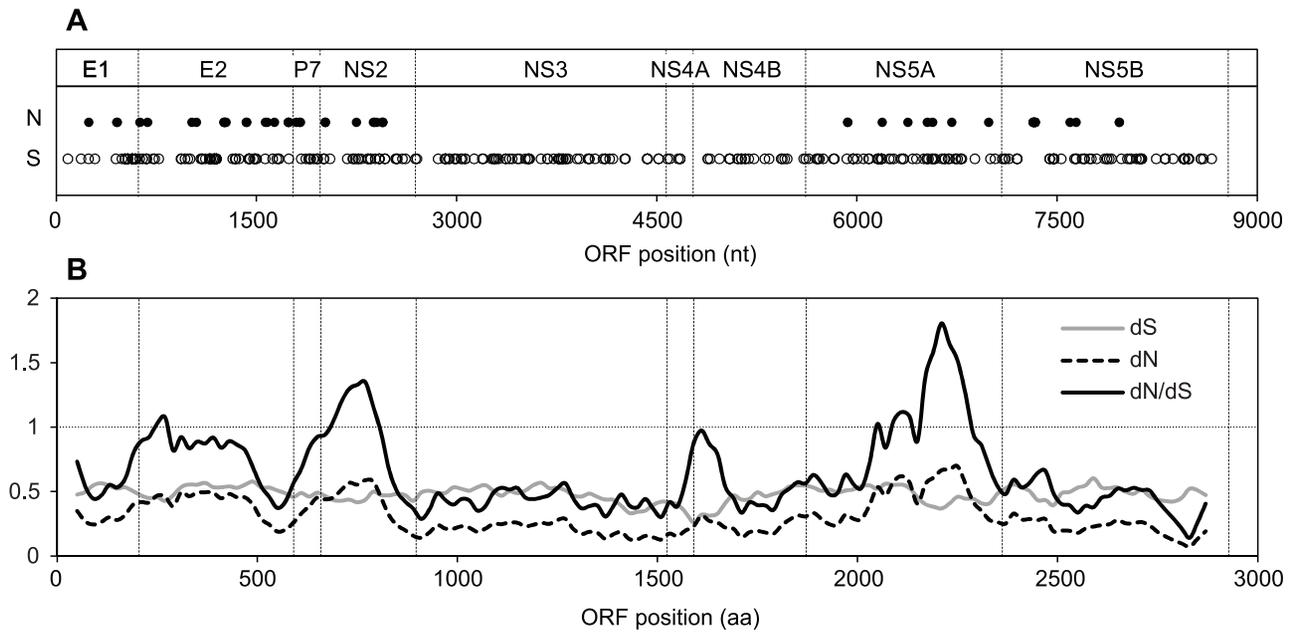
### Discussion

Here we report the discovery and characterization of novel pegiviruses in red colobus monkeys, red-tailed guenons and an olive baboon from western Uganda, an area that is a center of biodiversity [77] and a proposed “hotspot” for emerging infectious disease [28,78]. Our study describes the first pegiviruses detected in Old World monkeys and joins other emerging data demonstrating the impressive diversity of viruses infecting the endangered Kibale red colobus [29,30,79]. Overall, the high degree of conservation of genetic features among the Kibale SPgVs, HPgV and SPgV<sub>cpz</sub> suggests conservation of function among these related viruses. In fact, the distance between the Kibale SPgVs and HPgV is considerably less than the distance between HCV and its closest known relative, NPHV, whose natural host is horses [69,70].

This study supports the idea that pegivirus infection may be common among Old World and New World primates. To date, distinct SPgVs have been recovered from six New World monkey species: four tamarins (*Saguinus labiatus*, *S. mystax*, *S. nigricollis* and *S. oedipus*), an owl monkey (*Aotus trivirgatus*) and a marmoset (*C. jacchus*) [2,7,72]. Infection rates determined in wild-caught New World primates were generally high: 13 of 37 *S. mystax*, 7 of 9 *S. nigricollis*, 3 of 12 *S. labiatus*, 2 of 4 *S. oedipus*, 2 of 2 *C. jacchus*, and 6 of 12 *A. trivirgatus* [7]. Prior to our study, knowledge of pegivirus infections in Old World primates was limited to chimpanzees and humans: SPgV<sub>cpz</sub> viremia was previously detected in 3 of 39 wild and 6 of 235 captive chimpanzees [8,35], and the high prevalence of HPgV infection has been widely documented [18]. Here, a large proportion of the red colobus monkeys (28 of 60) and red-tailed guenons (5 of 12) sampled were viremic, suggesting that infections are common. In contrast, a single baboon (of 23 sampled) was positive, precluding similar inferences.

The Kibale SPgVs are monophyletic and share a most-recent common ancestor with HPgV and SPgV<sub>cpz</sub>, demonstrating the closely shared evolutionary history of these viruses in a geographic region where HPgV shows significant genetic variability [75]. Analysis of the partial-helicase phylogeny allows for comparison of the genetic diversity of most primate pegiviruses described to date across a well-studied coding sequence: the clades comprising pegiviruses that infect Old versus New World primate species now exhibit similar genetic diversity, while sequences of Kibale SPgV variants from animals of a given species were highly similar ( $\geq 94\%$  nt ID). Considering the much greater within-species diversity of known SPgV<sub>cpz</sub> variants (73% nt ID) and of several SPgV variants infecting New World monkeys (80 to 90% minimum nt ID within each species), the potential exists for significant undiscovered diversity in SPgVs infecting other populations of primates.

In contrast to the monophyletic sorting of Old World primate pegiviruses, the paraphyletic (i.e., multi-lineage) distribution of hepaci- and pegiviruses infecting bats and rodents is striking [12,14,15,61]. While this pattern may reflect the extensive



**Figure 4. Within- and between-host patterns of selection.** (A) Mapping of the distributions of synonymous (S) and nonsynonymous (N) single nucleotide polymorphisms along the ORF for 13 of 28 red colobus samples, for which coverage-depth was greater than 100 reads across most ( $\geq 96\%$ ) of the ORF. (B) Distribution of average substitution rates at non-synonymous (dN) and synonymous (dS) sites, and their ratio (dN/dS), along a sliding window (100 aa window, 20 aa step) for the comparison of selection pressures among the Kibale SPgVs (accession no. KF234523, KF234526 and KF234530).  
doi:10.1371/journal.pone.0098569.g004

sampling of these host taxa compared to primates, it also suggests extensive historical cross-species transmission of viruses within each distantly related genus [74]. For example, it is now clear that the New World SPgVs share a more recent common ancestor with a diverse array of bat pegiviruses than with Old World primate pegiviruses [12]. Furthermore, experimental infection of chimpanzees with HPgV [80] and laboratory passaging of New World SPgV isolates through different tamarin species [6] both demonstrate the capacity for extant pegiviruses to infect some closely related hosts.

Our assessments of within-host genetic diversity and between-host selection help to clarify the role of natural selection on the evolution of the Kibale SPgVs. In SPgV<sub>krc</sub>, we observed a relative abundance of low-frequency SNPs, and an even distribution of synonymous substitutions along the polyprotein. These observations suggest a significant and ongoing contribution of random mutations to SPgV within-host synonymous genetic diversity. On the other hand, we observed very low genetic diversity among the consensus SPgV<sub>krc</sub> sequences recovered from different animals during a 28-month period. These results indicate that purifying selection is likely acting to remove deleterious mutations in replicating viruses. If the evolution of SPgV<sub>krc</sub> were selectively neutral or strongly impacted by positive selection, we would expect to see greater genetic diversity among consensus sequences over this time period.

Among the different Kibale SPgVs, the dN/dS ratio, which estimates the relative rate of selected versus neutral changes across consensus sequences [81], varied across the polyprotein. We observed dN/dS values exceeding one in E2, NS2, and NS5A, suggesting positive selection on these genes. Interestingly, within individual SPgV samples – SPgV<sub>krc</sub> in particular – the distribution of nonsynonymous substitutions was similarly biased, with their prevalence in E2, P7, NS2 and NS5A suggesting greater tolerance and/or greater functional significance of nonsynonymous muta-

tions in these genes. Similarly, across NS3, the absence of nonsynonymous substitutions within individual samples combined with low and steady dN/dS ratios for each of the Kibale SPgVs signify the impacts purifying selection on this functionally important and constrained gene [82]. For NS5B we also observed low dN/dS ratios. This pattern is consistent with purifying selection, reflecting the region's critical role in RNA replication.

Persistent HPgV viremia is associated with prolonged survival and improved surrogate markers of disease progression in HIV-positive individuals [18,83,84]. Several potential mechanisms for the apparent protective effects of HPgV viremia have been identified, including direct antiviral effects; altered expression of cytokines, chemokines and HIV entry receptors; and modulation of host cell signaling pathways [18,19,27]. However, a great deal of uncertainty still surrounds HPgV mitigation of HIV pathogenesis, in part because no tractable animal model exists with which to study this phenomenon. It is noteworthy that natural SIV infection is present in the primates of western Uganda [79]. Whether SIV and SPgV interact in co-infected hosts is currently unclear. Despite early reports [85,86], HPgV has not been shown to infect macaques, the most developed and well-understood animal model of HIV pathogenesis, presumably because the barriers to host switching between humans and macaques are too great [3]. We speculate that these barriers might be more easily surmounted by the SPgVs described here, as macaques are much more closely related to the Old World monkey hosts of these viruses than they are to humans [87]. If SPgV variants productively infect macaques, SPgV-SIV coinfections may provide new avenues for understanding the mechanisms by which persistent HPgV infection antagonizes HIV pathogenesis.

## Materials and Methods

### Ethics Statement

The use of animal samples in this study followed the guidelines of the Weatherall Report on the use of non-human primates in research and was approved by the Uganda Wildlife Authority, the Uganda National Council for Science and Technology, and the University of Wisconsin Animal Care and Use Committee prior to initiation of the study.

### Study Site and Sample Collection

This study was conducted in Kibale National Park, western Uganda (0°13′–0°41′ N, 30°19′–30°32′ E), with prior approval from the Uganda National Council for Science and Technology and the Uganda Wildlife Authority. Kibale National Park is a semi-deciduous forest (795 km<sup>2</sup>) located near the Rwenzori Mountains and is notable for its biodiversity and density of primates [28,77]. Monkeys were immobilized and plasma was sampled as previously described [29]. Blood samples discussed in the current report were collected from January 2010 to June 2012, separated using centrifugation in a field laboratory, and frozen immediately in liquid nitrogen for storage and transport. All animal protocols received prior approval from the Uganda National Council for Science and Technology, the Uganda Wildlife Authority, and the University of Wisconsin Animal Care and Use Committee. All samples were shipped in accordance with international laws under Ugandan CITES permit #002290.

### RNA Extraction and Deep Sequencing

One ml blood plasma from each animal was filtered (0.45 μm) to remove residual host cells, and viral RNA was isolated using the Qiagen QIAamp MinElute virus spin kit (Qiagen, Hilden, Germany), omitting carrier RNA. The eluted RNA was treated with DNase I (DNA-free, Ambion, Austin, TX, USA), and double-stranded DNA was generated using the Superscript double-stranded cDNA Synthesis kit (Invitrogen, Carlsbad, CA, USA), primed with random hexamers. The DNA was purified using the Agencourt Ampure XP system (Beckman Coulter, Brea, CA, USA) and approximately 1 ng DNA was prepared for sequencing on an Illumina MiSeq (Illumina, San Diego, CA, USA) using the Nextera DNA sample preparation kit (Illumina, San Diego, CA, USA). Sequence data were analyzed using CLC Genomics Workbench version 5.5 (CLC bio, Aarhus, Denmark). Briefly, low-quality (CLC quality trimming limit = 0.001; phred quality score < 30) and short reads (< 100 bp) were removed and the remaining reads were subjected to *de novo* assembly. Assembled contiguous sequences (contigs) were queried against the GenBank database using the basic local alignment search tools blastn and blastx.

### Quantitative RT-PCR

Pegivirus titers were determined via TaqMan quantitative RT-PCR (qRT-PCR). RNA was extracted from 200 μL plasma using the Qiagen QIAamp MinElute virus spin kit (Qiagen, Hilden, Germany). Quantification was performed on a Lightcycler 480 real-time instrument (Roche, USA) using the SuperScript III Platinum One-Step Quantitative RT-PCR System kit (Invitrogen, Grand Island, NY). Pegiviruses detected in plasma from red colobus and red-tailed guenons were quantified using the following oligonucleotides, which were designed to anneal with all known Old World primate pegiviruses, including those described in the present work: forward 5′-TACGACGACTGCCCTACAC-3′ (750 nM); reverse 5′-TTTGGCCAGCTIACATCAGG-3′ (750 nM); probe 5′-FAM-CGCAGCCGTCGCTGCTGA-

CAT-BHQ1-3′ (150 nM). The pegivirus detected in plasma from an olive baboon was quantified using a separate assay, which was optimized to detect this virus: forward 5′-CGGTGTTTCATGGCAGGTAT-3′ (500 nM); reverse 5′-AAA-CACGCGGCTGTAAGT-3′ (500 nM); probe 5′-FAM-ATG-CACCCTGATGTAAGCTGGGCAA-BHQ1-3′ (100 nM). Assays were sensitive to between 10 and 100 genome copies per 20-μL qRT-PCR reaction. Standard cycling conditions for the SuperScript III qRT-PCR kit were used, with an annealing temperature of 58°C. Species-specific RNA standard curves were prepared from viral RNA that was (i) amplified with primers flanking each qRT-PCR amplicon (primer sequences are available upon request), (ii) cloned (Zero Blunt PCR Cloning Kit; Invitrogen), (iii) linearized (Hind III digestion; NEB, Ipswich, MA), (iv) transcribed (MEGAscript T7 kit; Ambion, Grand Island, NY), and (v) quantified fluorometrically (Qubit 2.0, Invitrogen, Grand Island, NY).

### Genome Characterization

Mature *Flaviviridae* structural proteins are cleaved from a single polyprotein by host signal peptidase and were distinguished here through manual [46] and *in silico* [47,48] signalase cleavage site prediction. N-glycosylation of envelope proteins was predicted using N-GlycoSite [88]. RNA secondary structure analysis of 5′- and 3′-untranslated regions (UTRs) was conducted using the RNAfold and RNAalifold algorithms, executed on the Vienna RNA Website server [38,39]. Full and targeted portions of UTR alignments were analyzed to obtain consensus structures of regions with significant structural conservation. The molecular weight of the putative ion channel protein, p7, was estimated using the pI/MW tool available on the ExPASy Bioinformatics Resource Portal [49]. Sequence-based analyses of NS5A to identify intrinsically disordered regions (IDRs) and sites within IDRs with capacity to undergo disorder-to-order transitions for binding interactions were conducted with the ANCHOR software package version 1.0 [63,64]. Linear motifs within IDRs potentially involved in protein-protein or substrate-protein interactions were identified by querying the Eukaryotic Linear Motif (ELM) database, using a conservative motif probability cutoff of 0.01 [65]. Sequences of the pegiviruses described in this manuscript were deposited in the Genbank database under accession numbers KF234499 to KF234530.

### Phylogenetic and Sequence Similarity Analyses

Complete coding sequences of 44 viruses available in GenBank (as of June 2013) were included in the phylogenetic analysis to capture the maximum diversity within known major clades of the hepaciviruses and pegiviruses. Codon-guided sequence alignments were generated via MAFFT and cleaned using Gblocks [89] using TranslatorX [90]. A Bayesian phylogenetic tree was constructed using MrBayes version 3.2.1 [23]. Only the first two codon positions of the nucleotide alignment were considered, since third codon positions demonstrated significant substitution saturation ( $p < 0.0001$ ), as determined using DAMBE version 5.3.38 [91]. Characters in the input alignment were partitioned by codon position, and model parameters were estimated independently from the data under default priors. Markov chains were run for 10 million generations, and robustness of phylogenetic groupings was assessed using posterior probability values calculated in MrBayes. The resulting majority rule consensus tree was displayed using FigTree version 1.3.1. The substitution model used in this analysis, GTR+I+Γ, was selected using jModelTest [92], and was based on the 5502-character alignment retained after Gblocks treatment and removal of third-codon-position nucleotides.

We also conducted phylogenetic analyses using a conserved 97-aa segment of the NS3 helicase gene commonly used for the taxonomy of hepaciviruses and pegiviruses (polyprotein positions 1221 to 1317 with reference to HPgV, NC\_001710) [7–9,72,73]. Following the codon-based alignment of 43 pegivirus and 25 hepacivirus sequences, we generated a neighbor-joining phylogeny (Poisson-corrected p-distance model, pairwise deletion, uniform rates) using MEGA5 (version 5.05) [93], with 5,000 bootstrap replicates of the data to assess the statistical confidence of phylogenetic groupings. Neighbor-joining methods based on amino-acid alignments are commonly used to classify novel hepaciviruses and pegiviruses when highly conserved genes (e.g., NS3 or NS5A) are targeted [14,61].

Recombination among Old World primate pegiviruses was assessed using GARD [94] and RDP3 [95]. Sequence identity (p-distance) was determined using MEGA5. Amino acid similarity between the novel and related primate pegiviruses was plotted across codon-aligned genomes by the sliding-window method implemented in SimPlot version 3.5.1 [96]. Because no recombinants were identified (see below), no sequences were excluded from phylogenetic analyses.

### Within-host Genetic Diversity

Single nucleotide polymorphism (SNP) analysis was performed using CLC's SNP analysis tool as previously described [61]. Stringent variant-calling criteria were used to ensure that only high-quality and high-coverage areas were considered in SNP calling (window = 7; maximum gap and mismatch count = 2; minimum central quality base = 30; minimum average quality for window = 25; minimum coverage = 100×; minimum variant frequency = 5%; SNP required in both forward and reverse reads). At the minimum accepted coverage of 100 high-quality reads and considering the requirement that SNPs be present in the forward and reverse direction, the theoretical detection threshold of this method is two SNPs, or 2%. Therefore, considering variants at a minimum frequency of 5% excludes any singleton observations (i.e., SNPs present on a single library fragment) and provides a conservative estimate of biologically relevant within-host genetic variation. This method also allows for direct comparisons with hepaciviruses and pegiviruses identified in previous studies [61,75].

### Between-host Selection

We aligned three representative pegivirus consensus sequences using via MAAFT, one from each positive monkey species (accession no. KF234523, KF234526 and KF234530), and estimated substitution rates at non-synonymous (dN) and synonymous (dS) sites, and their ratio (dN/dS), according to the method described by Nei and Gojobori [97] using the SNAP tool [98]. Values determined by SNAP for each amino acid in the alignment were then average along a sliding window (100 aa window, 20 aa step) for display.

### Supporting Information

**Figure S1 Observed viral titers for the three novel Kibale simian pegiviruses, SPgV<sub>krc</sub>, SPgV<sub>krtg</sub> and SPgV<sub>kbab</sub>, determined via quantitative TaqMan qRT-PCR.**

(EPS)

### References

- Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW, et al. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* 244: 359–362.

**Figure S2 Conserved RNA secondary structure in the 5'- and 3'-untranslated regions (UTRs) of Old World primate pegiviruses.**

Targeted portions of UTR alignments were analyzed using RNAalifold to obtain consensus structures and associated base-pair probabilities. Structures were numbered according to schemes established previously for pegivirus 5'- [36] and the 3'-UTRs [44]. Alignment positions along structures are listed for two reference sequences, HPgV strain Iowan (AF121950) and SPgV<sub>krc</sub>-RC01 (KF234505), respectively. The alignments analyzed comprised sequences from five red colobus (KF234505, KF234499, KF234523, KF234521, KF234520), one baboon (KF234530), two red-tailed guenon (KF234528, KF234529), one chimpanzee (AF070476) and five human (AF121950, U44402, KC618398, KC618400, KC618401) pegiviruses. Of note, sequence comprising the stem of stem-loop VII was absent in the 3'UTR of SPgV<sub>cpz</sub>.

(EPS)

**Figure S3 The distribution of synonymous substitutions, aggregated from 13 high-coverage SPgV<sub>krc</sub> samples, was proportional to the gene length across the ORF, suggesting a steady and essentially homogeneous contribution of random mutations to within-host synonymous genetic diversity.**

(EPS)

**Figure S4 The distribution of nonsynonymous substitutions, aggregated from 13 high-coverage SPgV<sub>krc</sub> samples, was poorly correlated to the gene length across the ORF, suggesting a differential impact of natural selection on within-host nonsynonymous genetic diversity.**

(EPS)

**Table S1 Taxa included in the Bayesian phylogenetic analysis (Fig. 2).**

(PDF)

**Table S2 Taxa included in the NS3 helicase neighbor-joining phylogenetic analysis (Fig. 3) in addition to those listed in Table S1. All taxa listed in Table S1 were included in this analysis.**

(PDF)

**Table S3 The distribution of synonymous and nonsynonymous substitutions, aggregated from 13 high-coverage SPgV<sub>krc</sub> samples, along the polyprotein versus gene length.**

(PDF)

### Acknowledgments

We are grateful to the Uganda Wildlife Authority, the Uganda National Council for Science and Technology, and Makerere University Biological Field Station for granting permission to conduct this research, J. Byaruhanga, P. Katurama, A. Nyamwija, J. Rusoke, A. Mbabazi, Dr. Dennis Twinomugisha, and Dr. Patrick Omeja for assistance in the field, and L. Kilby for assistance with permitting and logistics.

### Author Contributions

Conceived and designed the experiments: TLG DHO TCF CAC. Performed the experiments: SDS ML ALB TLG DH AT GW. Analyzed the data: SDS TLG DHO TCF. Wrote the paper: SDS ALB TLG DHO TCF.

3. Stapleton JT, Fong S, Muerhoff AS, Bukh J, Simmonds P (2011) The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae. *J Gen Virol* 92: 233–246.
4. Stapleton JT, Bukh J, Muerhoff AS, Fong S, Simmonds P (2012) Assignment of human, simian and bat pegiviruses (previously described as GBV-A, GBV-C, and GBV-D) as members of a new Genus (Pegivirus) within the Flaviviridae. Available: [http://talk.ictvonline.org/files/ictv\\_official\\_taxonomy\\_updates\\_since\\_the\\_8th\\_report/m/vertebrate-official/4486.aspx](http://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/vertebrate-official/4486.aspx). Accessed 15 July 2013.
5. Stapleton JT (2003) GB virus type C/Hepatitis G virus. *Semin Liver Dis* 23: 137–148.
6. Schaluder GG, Dawson GJ, Simons JN, Pilot-Matias TJ, Gutierrez RA, et al. (1995) Molecular and serologic analysis in the transmission of the GB hepatitis agents. *J Med Virol* 46: 81–90.
7. Bukh J, Appgar CL (1997) Five new or recently discovered (GBV-A) virus species are indigenous to New World monkeys and may constitute a separate genus of the Flaviviridae. *Virology* 229: 429–436.
8. Adams NJ, Prescott LE, Jarvis LM, Lewis JC, McClure MO, et al. (1998) Detection in chimpanzees of a novel flavivirus related to GB virus-C/hepatitis G virus. *J Gen Virol* 79: 1871–1877.
9. Birkenmeyer LG, Desai SM, Muerhoff AS, Leary TP, Simons JN, et al. (1998) Isolation of a GB virus-related genome from a chimpanzee. *J Med Virol* 56: 44–51.
10. Chandriani S, Skewes-Cox P, Zhong W, Ganem DE, Divers TJ, et al. (2013) Identification of a previously undescribed divergent virus from the Flaviviridae family in an outbreak of equine serum hepatitis. *Proc Natl Acad Sci* 110: E1407–E1415.
11. Kapoor A, Simmonds P, Cullen JM, Scheel T, Medina JL, et al. (2013) Identification of a pegivirus (GBV-like virus) that infects horses. *J Virol* 87: 7185–7190.
12. Quan P-L, Firth C, Conte JM, Williams SH, Zambrana-Torrel CM, et al. (2013) Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proc Natl Acad Sci* 110: 8194–8199.
13. Epstein JH, Quan P-L, Briese T, Street C, Jabado O, et al. (2010) Identification of GBV-D, a Novel GB-like Flavivirus from Old World Frugivorous Bats (*Pteropus giganteus*) in Bangladesh. *PLoS Pathog* 6: e1000972. doi:10.1371/journal.ppat.1000972.
14. Kapoor A, Simmonds P, Scheel TKH, Hjelle B, Cullen JM, et al. (2013) Identification of Rodent Homologs of Hepatitis C Virus and Pegiviruses. *mBio* 4: e00216–13.
15. Drexler JF, Corman VM, Müller MA, Lukashev AN, Gmyl A, et al. (2013) Evidence for Novel Hepaciviruses in Rodents. *PLoS Pathog* 9: e1003438. doi:10.1371/journal.ppat.1003438.
16. Muerhoff AS, Leary TP, Sathar MA, Dawson GJ, Desai SM (2005) African origin of GB virus C determined by phylogenetic analysis of a complete genotype 5 genome from South Africa. *J Gen Virol* 86: 1729–1735.
17. Hitzler WE, Runkel S (2004) Prevalence, persistence and liver enzyme levels of HGV RNA-positive blood donors determined by large-scale screening and transmission by blood components. *Clin Lab* 50: 25–31.
18. Bhattarai N, Stapleton JT (2012) GB virus C: the good boy virus? *Trends Microbiol* 20: 124–130.
19. Mohr EL, Stapleton JT (2009) GB virus type C interactions with HIV: the role of envelope glycoproteins. *J Viral Hepat* 16: 757–768.
20. Giret MTM, Miraglia JL, Sucupira MCA, Nishiya A, Levi JE, et al. (2011) Prevalence, Incidence Density, and Genotype Distribution of GB Virus C Infection in a Cohort of Recently HIV-1-Infected Subjects in Sao Paulo, Brazil. *PLoS ONE* 6: e18407. doi:10.1371/journal.pone.0018407.
21. Rey D, Vidinic-Moullarde J, Meyer P, Schmitt C, Fritsch S, et al. (2000) High prevalence of GB virus C/hepatitis G virus RNA and antibodies in patients infected with human immunodeficiency virus type 1. *Eur J Clin Microbiol Infect Dis* Off Publ Eur Soc Clin Microbiol 19: 721–724.
22. Schwarze-Zander C, Blackard JT, Zheng H, Addo MM, Lin W, et al. (2006) GB virus C (GBV-C) infection in hepatitis C virus (HCV)/HIV-coinfected patients receiving HCV treatment: importance of the GBV-C genotype. *J Infect Dis* 194: 410–419.
23. Bhanich Supapol W, Remis RS, Raboud J, Millson M, Tappero J, et al. (2011) Prevalence and correlates of GB virus C infection in HIV-infected and HIV-uninfected pregnant women in Bangkok, Thailand. *J Med Virol* 83: 33–44.
24. Charrel RN, De Micco P, de Lamballerie X (1999) Phylogenetic analysis of GB viruses A and C: evidence for cospeciation between virus isolates and their primate hosts. *J Gen Virol* 80: 2329–2335.
25. Romano CM, Zanotto PM de A, Holmes EC (2008) Bayesian coalescent analysis reveals a high rate of molecular evolution in GB virus C. *J Mol Evol* 66: 292–297.
26. Pavesi A (2001) Origin and evolution of GBV-C/hepatitis G virus and relationships with ancient human migrations. *J Mol Evol* 53: 104–113.
27. Bhattarai N, McLinden JH, Xiang J, Landay AL, Chivero ET, et al. (2013) GB Virus C Particles Inhibit T Cell Activation via Envelope E2 Protein-Mediated Inhibition of TCR Signaling. *J Immunol Baltim Md* 1950 190: 6351–6359.
28. Goldberg TL, Paige SB, Chapman CA (2012) The Kibale EcoHealth Project: exploring connections among human health, animal health, and landscape dynamics in western Uganda. In: Aguirre AA, Ostfeld RS, Daszak P (ed.), *New directions in conservation medicine: Applied cases of ecological health*. Oxford, New York: Oxford University Press. 452–465.
29. Lauck M, Hyeroba D, Tumukunde A, Weny G, Lank SM, et al. (2011) Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing. *PLoS One* 6: e19056. doi:10.1371/journal.pone.0019056.
30. Goldberg TL, Chapman CA, Cameron K, Saj T, Karesh WB, et al. (2008) Serologic evidence for novel poxvirus in endangered red Colobus monkeys, Western Uganda. *Emerg Infect Dis* 14: 801–803.
31. Salyer SJ, Gillespie TR, Rwego IB, Chapman CA, Goldberg TL (2012) Epidemiology and Molecular Relationships of *Cryptosporidium* spp. in People, Primates, and Livestock from Western Uganda. *PLoS Negl Trop Dis* 6: e1597. doi:10.1371/journal.pntd.0001597.
32. Johnston AR, Gillespie TR, Rwego IB, McLachlan TLT, Kent AD, et al. (2010) Molecular epidemiology of cross-species *Giardia duodenalis* transmission in western Uganda. *PLoS Negl Trop Dis* 4: e683. doi:10.1371/journal.pntd.0000683.
33. Thurber MI, Ghai RR, Hyeroba D, Weny G, Tumukunde A, et al. (2013) Co-infection and cross-species transmission of divergent *Hepaticystis* lineages in a wild African primate community. *Int J Parasitol* 43: 613–619. doi:10.1016/j.ijpara.2013.03.002.
34. Goldberg TL, Gillespie TR, Rwego IB, Estoff EL, Chapman CA (2008) Forest Fragmentation as Cause of Bacterial Transmission among Nonhuman Primates, Humans, and Livestock, Uganda. *Emerg Infect Dis* 14: 1375–1382. doi:10.3201/eid1409.071196.
35. Mohr EL, Murthy KK, McLinden JH, Xiang J, Stapleton JT (2011) The natural history of non-human GB virus C in captive chimpanzees. *J Gen Virol* 92: 91–100.
36. Simons JN, Desai SM, Schultz DE, Lemon SM, Mushahwar IK (1996) Translation initiation in GB viruses A and C: evidence for internal ribosome entry and implications for genome organization. *J Virol* 70: 6126–6135.
37. Turner C, Witwer C, Hofacker IL, Stadler PF (2004) Conserved RNA secondary structures in Flaviviridae genomes. *J Gen Virol* 85: 1113–1124.
38. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF (2008) RNAfold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9: 474.
39. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL (2008) The Vienna RNA Website. *Nucleic Acids Res* 36: W70–W74.
40. Chang K-S, Luo G (2006) The polypyrimidine tract-binding protein (PTB) is required for efficient replication of hepatitis C virus (HCV) RNA. *Virus Res* 115: 1–8.
41. Yanagi M, St Claire M, Emerson SU, Purcell RH, Bukh J (1999) In vivo analysis of the 3' untranslated region of the hepatitis C virus after in vitro mutagenesis of an infectious cDNA clone. *Proc Natl Acad Sci* 96: 2291–2295.
42. Wang L, Jeng K-S, Lai MMC (2011) Poly(C)-binding protein 2 interacts with sequences required for viral replication in the hepatitis C virus (HCV) 5' untranslated region and directs HCV RNA replication through circularizing the viral genome. *J Virol* 85: 7954–7964.
43. Holcik M, Liebhaber SA (1997) Four highly stable eukaryotic mRNAs assemble 3' untranslated region RNA-protein complexes sharing cis and trans components. *Proc Natl Acad Sci* 94: 2410–2414.
44. Cuceanu NM, Tuplin A, Simmonds P (2001) Evolutionarily conserved RNA secondary structures in coding and non-coding sequences at the 3' end of the hepatitis G virus/GB-virus C genome. *J Gen Virol* 82: 713–722.
45. Mandl CW, Heinz FX, Puchhammer-Stöckl E, Kunz C (1991) Sequencing the termini of capped viral RNA by 5'-3' ligation and PCR. *BioTechniques* 10: 484–486.
46. Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 10: 1–6.
47. Reynolds SM, Kall L, Riffle ME, Bilmes JA, Noble WS (2008) Transmembrane Topology and Signal Peptide Prediction Using Dynamic Bayesian Networks. *PLoS Comput Biol* 4: e1000213. doi:10.1371/journal.pcbi.1000213.
48. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8: 785–786.
49. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, et al. (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM (ed.), *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press. 571–607.
50. Erker JC, Desai SM, Leary TP, Chalmers ML, Montes CC, et al. (1998) Genomic analysis of two GB virus A variants isolated from captive monkeys. *J Gen Virol* 79: 41–45.
51. Welbourn S, Pause A (2006) HCV NS2/3 Protease. Available: <http://www.ncbi.nlm.nih.gov/books/NBK1622/>. Accessed 2013 Mar 14.
52. Lin C (2006) HCV NS3-4A Serine Protease. Available: <http://www.ncbi.nlm.nih.gov/books/NBK1623/>. Accessed 2013 Mar 14.
53. Belyaev AS, Chong S, Novikov A, Kongpachith A, Masiarz FR, et al. (1998) Hepatitis G Virus Encodes Protease Activities Which Can Effect Processing of the Virus Putative Nonstructural Proteins. *J Virol* 72: 868–872.
54. Lee VJ, Kang H, Rho S-H, Eom SH, Park WJ (2001) Assessment of Substrate Specificity of Hepatitis G Virus NS3 Protease by a Genetic Method. *Biochem Biophys Res Commun* 286: 171–175.
55. George SL, Varmaz D, Tavis JE, Chowdhury A (2012) The GB Virus C (GBV-C) NS3 Serine Protease Inhibits HIV-1 Replication in a CD4+ T Lymphocyte

- Cell Line without Decreasing HIV Receptor Expression. *PLoS ONE* 7: e30653. doi:10.1371/journal.pone.0030653.
56. Koonin EV, Dolja VV (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol* 28: 375–430.
  57. Koonin EV (1991) The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J Gen Virol* 72: 2197–2206.
  58. Tellinghuisen TL, Paulson MS, Rice CM (2006) The NS5A Protein of Bovine Viral Diarrhea Virus Contains an Essential Zinc-Binding Site Similar to That of the Hepatitis C Virus NS5A Protein. *J Virol* 80: 7450–7458.
  59. Feuerstein S, Solyom Z, Aladag A, Favier A, Schwarten M, et al. (2012) Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J Mol Biol* 420: 310–323.
  60. Gupta G, Qin H, Song J (2012) Intrinsically Unstructured Domain 3 of Hepatitis C Virus NS5A Forms a “Fuzzy Complex” with VAPB-MSP Domain Which Carries ALS-Causing Mutations. *PLoS ONE* 7: e39261. doi:10.1371/journal.pone.0039261.
  61. Lauck M, Sibley SD, Lara J, Purdy MA, Khudyakov Y, et al. (2013) A novel hepacivirus with an unusually long and intrinsically disordered NS5A protein in a wild Old World primate. *J. Virol.* 87: 8971–8981.
  62. Dosztányi Z, Csizmek V, Tompa P, Simon I (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21: 3433–3434.
  63. Dosztányi Z, Mészáros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinforma Oxf Engl* 25: 2745–2746.
  64. Mészáros B, Simon I, Dosztányi Z (2009) Prediction of Protein Binding Regions in Disordered Proteins. *PLoS Comput Biol* 5: e1000376. doi:10.1371/journal.pcbi.1000376.
  65. Dinkel H, Michael S, Weatheritt RJ, Davey NE, Van Roey K, et al. (2012) ELM—the database of eukaryotic linear motifs. *Nucleic Acids Res* 40: D242–251.
  66. Bouwmeester T, Bauch A, Ruffner H, Angrand P-O, Bergamini G, et al. (2004) A physical and functional map of the human TNF- $\alpha$ /NF- $\kappa$ B signal transduction pathway. *Nat Cell Biol* 6: 97–105.
  67. Takeuchi M, Rothe M, Goeddel DV (1996) Anatomy of TRAF2 distinct domains for nuclear factor- $\kappa$ B activation and association with tumor necrosis factor signaling proteins. *J Biol Chem* 271: 19935–19942.
  68. Takahashi K, Hijikata M, Hino K, Mishiro S (1997) Entire polyprotein-ORF sequences of Japanese GBV-C/HGV isolates: implications for new genotypes. *Hepatology* 25: 139–148.
  69. Lyons S, Kapoor A, Sharp C, Schneider BS, Wolfe ND, et al. (2012) Nonprimate Hepaciviruses in Domestic Horses, United Kingdom. *Emerg Infect Dis* 18: 1976–1982.
  70. Burbelo PD, Dubovi EJ, Simmonds P, Medina JL, Henriquez JA, et al. (2012) Serology-enabled discovery of genetically diverse hepaciviruses in a new host. *J Virol* 86: 6171–6178.
  71. Perelman P, Johnson WE, Roos C, Seuánez HN, Horvath JE, et al. (2011) A Molecular Phylogeny of Living Primates. *PLoS Genet* 7: e1001342. doi:10.1371/journal.pgen.1001342.
  72. Leary TP, Desai SM, Yamaguchi J, Chalmers ML, Schlauder GG, et al. (1996) Species-specific variants of GB virus A in captive monkeys. *J Virol* 70: 9028–9030.
  73. Leary TP, Desai SM, Erker JC, Mushahwar IK (1997) The sequence and genomic organization of a GB virus A variant isolated from captive tamarins. *J Gen Virol* 78: 2307–2313.
  74. Pybus OG, Gray RR (2013) Virology: The virus whose family expanded. *Nature* 498: 310–311.
  75. Ghai R, Sibley SD, Lauck M, Dinis J, Bailey AL, et al. (2013) Deep sequencing identifies two genotypes and high viral genetic diversity of human pegivirus (GB virus C) among rural Ugandan patients. *J Gen Virol* 94: 2670–2678.
  76. Lauck M, Alvarado-Mora MV, Becker EA, Bhattacharya D, Striker R, et al. (2012) Analysis of Hepatitis C Virus Intra-Host Diversity Across The Coding Region by Ultra-Deep Pyrosequencing. *J Virol* 86: 3952–3960.
  77. Plumptre AJ, Davenport TRB, Behangana M, Kityo R, Eilu G, et al. (2007) The biodiversity of the Albertine Rift. *Biol Conserv* 134: 178–194.
  78. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, et al. (2008) Global trends in emerging infectious diseases. *Nature* 451: 990–993.
  79. Goldberg TL, Sintasath DM, Chapman CA, Cameron KM, Karesh WB, et al. (2009) Coinfection of Ugandan red colobus (*Procolobus [Ptilocolobus] rufomitratus tephrosceles*) with novel, divergent delta-, lenti-, and spumaretroviruses. *J Virol* 83: 11318–11329.
  80. Bukh J, Kim JP, Govindarajan S, Apgar CL, Fong SK, et al. (1998) Experimental infection of chimpanzees with hepatitis G virus and genetic analysis of the virus. *J Infect Dis* 177: 855–862.
  81. Kryazhinskiy S, Plotkin JB (2008) The Population Genetics of dN/dS. *PLoS Genet* 4: e1000304. doi:10.1371/journal.pgen.1000304.
  82. Parreira R, Branco C, Piedade J, Esteves A (2012) GB virus C (GBV-C) evolutionary patterns revealed by analyses of reference genomes, E2 and NS5B sequences amplified from viral strains circulating in the Lisbon area (Portugal). *Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis* 12: 86–93.
  83. Tillmann HL, Heiken H, Knapik-Botor A, Heringlake S, Ockenga J, et al. (2001) Infection with GB virus C and reduced mortality among HIV-infected patients. *N Engl J Med* 345: 715–724.
  84. Heringlake S, Ockenga J, Tillmann HL, Trautwein C, Meissner D, et al. (1998) GB virus C/hepatitis G virus infection: a favorable prognostic factor in human immunodeficiency virus-infected patients? *J Infect Dis* 177: 1723–1726.
  85. Cheng Y, Zhang W, Li J, Li B a, Zhao J m, et al. (2000) Serological and histological findings in infection and transmission of GBV-C/HGV to macaques. *J Med Virol* 60: 28–33.
  86. Ren H, Zhu F-L, Cao M-M, Wen X-Y, Zhao P, et al. (2005) Hepatitis G virus genomic RNA is pathogenic to *Macaca mulatta*. *World J Gastroenterol WJG* 11: 970–975.
  87. Longdon B, Hadfield JD, Webster CL, Obbard DJ, Jiggins FM (2011) Host Phylogeny Determines Viral Persistence and Replication in Novel Hosts. *PLoS Pathog* 7: e1002260. doi:10.1371/journal.ppat.1002260.
  88. Zhang M, Gaschen B, Blay W, Foley B, Haigwood N, et al. (2004) Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* 14: 1229–1246.
  89. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56: 564–577.
  90. Abascal F, Zardoya R, Telford MJ (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res*. Available: <http://nar.oxfordjournals.org/content/early/2010/04/30/nar.gkq291>. Accessed 17 July 2013.
  91. Xia X (2013) DAMBE5: A comprehensive software package for data analysis in molecular biology and evolution. *Mol Biol Evol* 38: W7–W13.
  92. Guindon S, Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst Biol* 52: 696–704.
  93. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 28: 2731–2739.
  94. Delpont W, Poon AFY, Frost SDW, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinforma Oxf Engl* 26: 2455–2457.
  95. Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26: 2462–2463.
  96. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, et al. (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73: 152–160.
  97. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3: 418–426.
  98. Korber B (2000) HIV Signature and Sequence Variation Analysis. In: Rodrigo AG, Learn GH. (eds.), *Computational Analysis of HIV Molecular Sequences*. Dordrecht, Netherlands: Kluwer Academic Publishers. 55–72.