Discovery of Novel Herpes Simplexviruses in Wild Gorillas, Bonobos, and Chimpanzees Supports Zoonotic Origin of HSV-2

Joel O. Wertheim,*,[1] Reilly Hostager,[1] Diane Ryu,[2,3] Kevin Merkel,[2,3] Samuel Angedakin,[4] Mimi Arandjelovic,[4] Emmanuel Ayuk Ayimisin,[4] Fred Babweteera,[5] Mattia Bessone,[4,6] Kathryn J. Brun-Jeffery,[7] Paula Dieguez,[4] Winnie Eckardt,[8,9] Barbara Fruth,[6,10] Ilka Herbinger,[11] Sorrel Jones,[4,12] Hjalmar Kuehl,[4,13] Kevin E. Langergraber,[14,15] Kevin Lee,[4,14,15] Nadege F. Madinda,[3,4] Sonja Metzger,[3,4] Lucy Jayne Ormsby,[4] Martha M. Robbins,[4] Volker Sommer,[16,17] Tara Stoinski,[8,9] Erin G. Wessling,[4,18] Roman M. Wittig,[4] Yisa Ginath Yuh,[†,4] Fabian H. Leendertz,[3] and Sébastien Calvignac-Spencer (iD)*,[2,3]

[1]Department of Medicine, University of California San Diego, San Diego, CA, USA
[2]Viral Evolution, Robert Koch Institute, Berlin, Germany
[3]Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany
[4]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[5]Budongo Conservation Field Station, Masindi, Uganda
[6]School of Biological & Environmental Sciences, Liverpool John Moores University, Liverpool, United Kingdom
[7]Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom
[8]Dian Fossey Gorilla Fund International, Atlanta, GA, USA
[9]Department of Environmental Sciences and Program in Population Biology, Ecology and Evolution, Emory University, Druid Hills, GA, USA
[10]Centre for Research and Conservation, Royal Zoological Society of Antwerp, Antwerp, Belgium
[11]WWF Germany, Berlin, Germany
[12]Royal Society for the Protection of Birds, Centre for Conservation Science, Cambridge, United Kingdom
[13]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
[14]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA
[15]Institute of Human Origins, Arizona State University, Tempe, AZ, USA
[16]Department of Anthropology, University College London, London, United Kingdom
[17]Gashaka Primate Project, Serti/Taraba, Nigeria
[18]Human Evolutionary Biology, Harvard University, Cambridge, MA, USA
[†]Present address: University of Concordia, Montreal, Canada
*Corresponding authors: E-mails: jwertheim@health.ucsd.edu; calvignacs@rki.de.
Associate editor: Tal Pupko

Abstract

Viruses closely related to human pathogens can reveal the origins of human infectious diseases. Human herpes simplexvirus type 1 (HSV-1) and type 2 (HSV-2) are hypothesized to have arisen via host-virus codivergence and cross-species transmission. We report the discovery of novel herpes simplexviruses during a large-scale screening of fecal samples from wild gorillas, bonobos, and chimpanzees. Phylogenetic analysis indicates that, contrary to expectation, simplexviruses from these African apes are all more closely related to HSV-2 than to HSV-1. Molecular clock-based hypothesis testing suggests the divergence between HSV-1 and the African great ape simplexviruses likely represents a codivergence event between humans and gorillas. The simplexviruses infecting African great apes subsequently experienced multiple cross-species transmission events over the past 3 My, the most recent of which occurred between humans and bonobos around 1 Ma. These findings revise our understanding of the origins of human herpes simplexviruses and suggest that HSV-2 is one of the earliest zoonotic pathogens.

Key words: phylogenetics, zoonosis, herpesvirus, molecular clock, ape.

Introduction

The origins of pathogens that infect humans can be revealed by the characterization of closely related microorganisms in wild and domesticated animals. Phylogenetic analysis has revealed the source of zoonoses both recent (e.g., HIV/AIDS, Ebola, influenza, MERS, SARS, COVID-19; Leroy et al. 2005; Hon et al. 2008; Smith et al. 2009; Sharp and Hahn 2011; Dudas et al. 2018; Nelson and Worobey 2018; Lu et al. 2020) and ancient (e.g., measles, smallpox, bubonic plague; Wertheim and Kosakovsky Pond 2011; Wagner et al. 2014; Duggan et al. 2016; Dux et al. 2020) alike. In contrast to these zoonoses, some pathogens may have been infecting our primate ancestors and have since evolved into human-specific forms via codivergence (e.g., varicella zoster virus and human mastadenovirus species C; Weiss 2007; Grose 2012; Hoppe et al. 2015). The human herpes simplexviruses (HSVs) appear to have arisen from both of these mechanisms: codivergence and zoonosis (Luebcke et al. 2006; Weiss 2007; Severini et al. 2013; Wertheim et al. 2014).

HSVs belong to a diverse family of viruses with double-stranded DNA genomes (Herpesviridae), whose members infect a broad range of vertebrates and invertebrates, reflecting a remarkably ancient association of these viruses with their hosts (McGeoch et al. 1995; Davison et al. 2009). Within this family, HSVs are part of a subfamily (Alphaherpesvirinae) and genus (Simplexvirus) within which codivergence events are very frequent. This pattern is exemplified by primate-infecting simplexviruses whose phylogeny recapitulates that of Platyrrhini (New World monkeys), Cercopithecidae (Old World monkeys), and Homininae (African great apes, including humans) (Eberle and Black 1993). Accordingly, most primate species have only been associated with a single simplexvirus. Humans are a striking exception in that they are infected with two: HSV-1 (Human alphaherpesvirus 1) and HSV-2 (Human alphaherpesvirus 2). Cross-sectional estimates indicate that HSV-1 infects over two-thirds of adults worldwide, whereas HSV-2 has a lower prevalence around 11% (Smith and Robinson 2002). HSV-1 is predominantly associated with oral transmission early in life, and HSV-2 is more commonly associated with sexual transmission; however, the disease presentation by these viruses is indistinguishable and both viruses can transmit through oral and genital routes. Both viruses can experience long periods of latency between infection and transmission, which when coupled with robust DNA repair mechanisms, may explain their exceptionally slow rate of evolution: around 10^{-8} substitutions/site/year (Norberg et al. 2011).

HSV-1 and HSV-2 are evolutionarily very closely related. They are not, however, each other's closest relatives. HSV-2 is more closely related to the chimpanzee herpes virus (ChHV; Panine alphaherpesvirus 3), which was discovered in a captive chimpanzee (Pan troglodytes), than either of these viruses is to HSV-1 (Luebcke et al. 2006; Severini et al. 2013). Two alternate scenarios have been articulated to account for this pattern, which both assume that ChHV is a codiverged simplexvirus infecting (wild) chimpanzees and suggest a potential zoonotic origin for one of the HSVs. Pairwise genetic distance

analysis suggested that HSV-1 was of zoonotic origin from a hypothetical orangutan simplexvirus ancestor and that HSV-2 had codiverged with ChHV (Luebcke et al. 2006; Severini et al. 2013). In contrast, we previously suggested that HSV-2 was of recent zoonotic origin from the ancestor of modern P. troglodytes and that HSV-1 has its origins in codivergence (Wertheim et al. 2014). Both of these scenarios assume there exist, or once existed (Johnson et al. 2003), simplexviruses infecting great apes (Hominidae) other than chimpanzees: bonobo (Pan paniscus), gorilla (Gorilla spp.), and orangutan (Pongo spp.). Great apes can be infected with HSV-1 and HSV-2 (Eberle and Hilliard 1989); however, the presence of extant nonhuman Homininae simplexviruses is supported by serology in wild Gorilla beringei beringei populations that demonstrates intermediate ELISA reactivity to both HSV-1 and HSV-2 antigens, with preferential reactivity toward HSV-2 (Eberle 1992).

Both hypothesized scenarios about the origins of HSV-1 and HSV-2 make specific predictions about the phylogenetic placement of uncharacterized viruses. The HSV-1 zoonotic origin hypothesis assumes the split between HSV-2 and ChHV represents a codivergence event and predicts that a gorilla simplexvirus would lie basal to HSV-2 and ChHV, but that HSV-1 would lie basal to these three viruses. In contrast, the HSV-2 zoonotic origin hypothesis assumes the split between HSV-1 and ChHV represents a codivergence event and predicts that a gorilla simplexvirus would lie basal to HSV-1, HSV-2, and ChHV. Here, we report the results of a large-scale molecular screening of fecal samples from wild animals representing all African great ape species, including seven of nine African great ape subspecies. This screening allowed us to confirm ChHV infection in wild West African chimpanzees (P. troglodytes verus) as well as detect simplexviruses infecting wild mountain gorillas (G. beringei beringei) and bonobos (P. paniscus). We determined partial gene sequences from virus infecting these three host species, and we assembled a partial genome sequence for the simplexvirus infecting mountain gorillas. Phylogenetic and molecular clock analyses suggest a complicated history of cross-species transmission of simplexviruses among Homininae species.

Results

Discovery and Sequencing of Wild Homininae Simplexviruses

We screened a large set of African great ape feces, identifying eight positive samples with a nested PCR system targeting a short fragment of the glycoprotein B (gB) coding sequence. We found a single positive sample in the set of feces from a pregnant Western chimpanzee (P. troglodytes verus), two positive samples from one site and one from another site, representing at least two bonobos (P. paniscus), and four positive samples from a pregnant ($n = 3$) and another lactating ($n = 1$) mountain gorilla (G. beringei beringei); for all these samples, we sequenced this 124-bp gB fragment (UL27). We attempted to obtain more sequence information by using additional semi-nested and nested PCR systems targeting gB and UL53 but were only successful with samples from
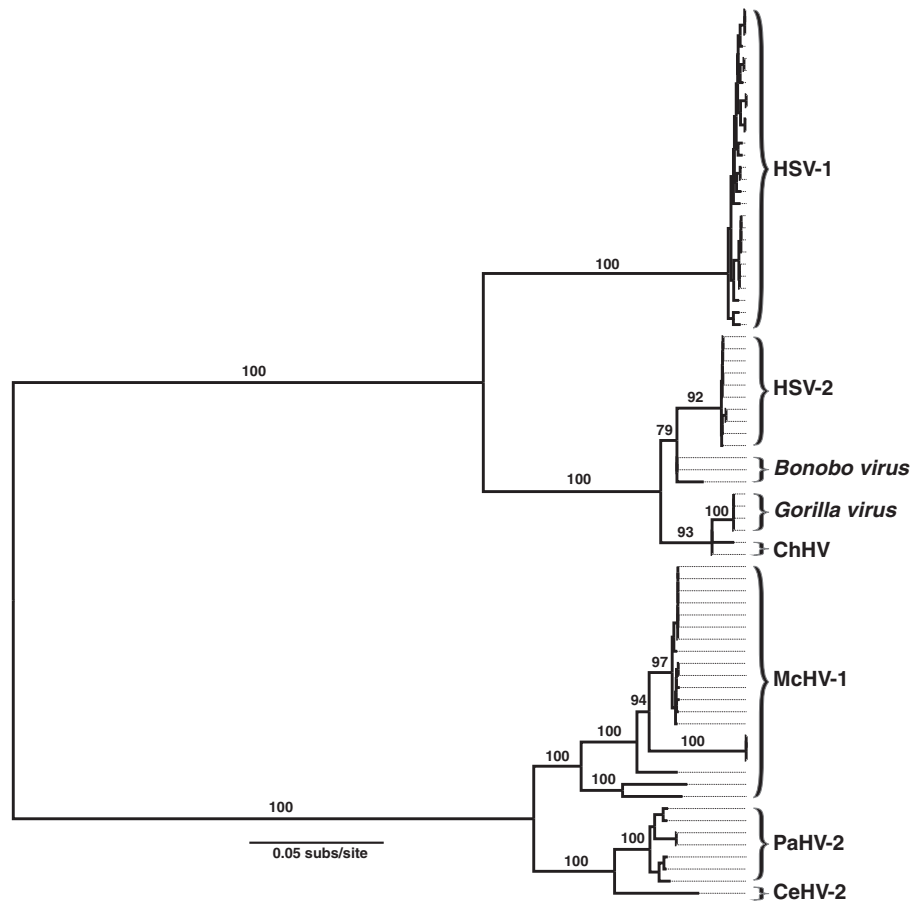
Fig. 1. Maximum likelihood phylogenetic tree of Cercopithecidae and Homininae herpes simplexviruses. The tree includes 74 taxa and was inferred using an alignment of 29 coding genes totaling 43,797 bp in length. Node support from ultrafast bootstrap analysis is denoted on major branches separating virus infecting different primate species. HSV-1 and HSV-2, human herpes simplexviruses 1 and 2; ChHV, chimpanzee herpes virus; McHV-1, macacine herpes virus 1; PaHV-2, Baboon herpes virus 2; CeHV-2, Cercopithecus herpes virus 2. Tree with strain labels and support values is available on Data Dryad.

pregnant mountain gorillas. Sequencing these amplicons allowed us to sequence 410 and 272 bp of the gB and UL53 coding sequences, respectively.

To collect genome-wide sequence information, we also attempted to enrich high-throughput sequencing libraries generated from all positive samples in simplexvirus sequences using hybridization capture. We generated a total of 60,741,470 reads (2,163,068–22,702,256 reads per sample, after filtering). Only the libraries generated from the pregnant mountain gorilla contained enough simplexvirus reads to allow for the reconstruction of a significant fraction of the genome. We were able to unambiguously characterize 103,446 positions of the according viral genome, representing 74.8% of its unique and internal repeat regions, from which we identified 58 coding sequences homologous (and syntenic) to those found in HSV-2.

Partial Genome Maximum Likelihood Phylogeny

The herpes simplexvirus maximum likelihood phylogeny including previously described catarrhine simplexviruses and the newly described viruses supports a general pattern of viral codivergence with their host species, with strong node support between major lineages (fig. 1 and see Supplementary Material online). All viruses from Homininae are reciprocally monophyletic with respect to viruses from Cercopithecidae (ultrafast bootstrap support [UF] = 100). Each Homininae species appears to have a distinct simplexvirus, with the known exception of HSV-1 and HSV-2 in *Homo sapiens*. However, these Homininae simplexviruses do not recapitulate their host phylogeny (fig. 2A). We caution against overinterpretation of the tanglegrams (de Vienne 2019) and note the differences in structure between the host and virus trees. Specifically, the nonhuman Homininae simplexviruses are more closely related to HSV-2 than they are to HSV-1 (UF = 100). The closest relative of the novel gorilla simplexvirus is ChHV (UF = 93), and the closet relative of HSV-2 is the novel bonobo simplexvirus (UF = 79). In contrast, *P. troglodytes* and *P. paniscus* are most closely related and both more closely related to *H. sapiens* than they are to *G. gorilla*. The identical topology is observed across the phylogeny if we exclude the shorter sequences from *G. gorilla*, *P. troglodytes*, *P. paniscus*, and *Macaca fuscata* from the alignment prior to phylogenetic inference. A phylogeny inferred from all taxa restricted to
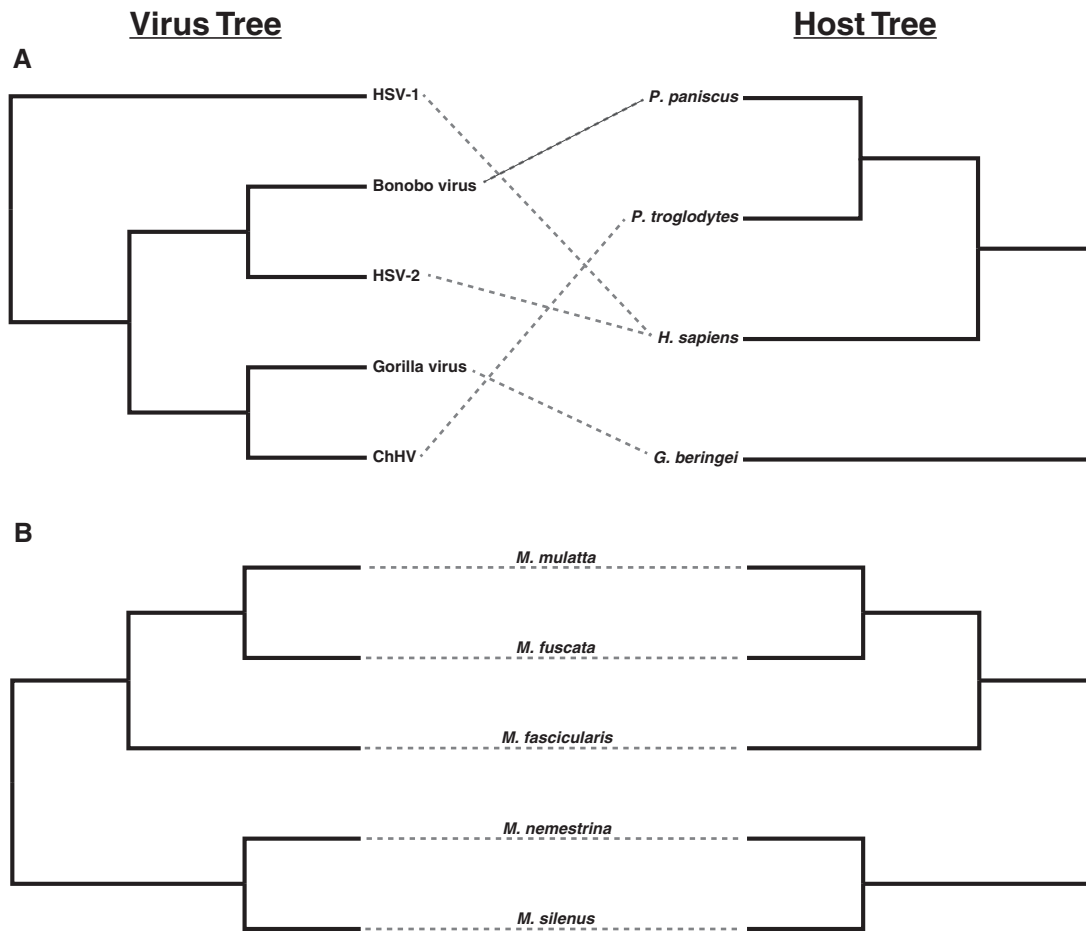
FIG. 2. Phylogenetic concordance between host and simplexvirus phylogenies. (*A*) Relationship between Homininae simplexviruses (right) and their hosts (left). (*B*) Relationship between *Macaca* simplexviruses and their hosts. All *Macaca* viruses are named McHV-1, so the trees are labeled only with host names. Virus trees are shown as cladograms based on the maximum likelihood phylogeny depicted in figure 1.

the glycoprotein B (UL27) gene in which many of these gene fragments reside is highly similar to the partial genome inference, albeit with lower UF (supplementary fig. 1, Supplementary Material online). We did not detect evidence of positive, diversifying selection in UL27 (supplementary fig. 1, Supplementary Material online), suggesting that this lower topological support is due to its shorter length, rather than biased inference due to positive selection.

The close phylogenetic relatedness of HSV-2 to the novel bonobo simplexvirus is supported by only a single nonsynonymous nucleotide substitution across the 124 nucleotide novel bonobo simplexvirus sequence: cytosine (isoleucine in all basal lineages) to guanine (methionine in HSV-2 and the *P. paniscus* simplexvirus) (supplementary fig. 2, Supplementary Material online). Regardless of the specific placement of the novel bonobo simplexvirus, the remaining phylogenetic relationships are consistent with multiple cross-species transmission of simplexvirus among Homininae. If the ancestor of this clade infected the immediate ancestor of gorillas, then the primate precursors to both humans and chimpanzees would have acquired this virus via cross-species transmission; however, the order and direction of these events remain ambiguous.

Two major genera within the Cercopithecidae family, *Papio* and *Macaca*, are each represented by a distinct viral clade: PaHV-2 (*Papiine alphaherpesvirus 2*, formerly known as HVP2) and McHV-1 (*Macacine alphaherpesvirus 1*, formerly known as B virus). The placement of CeHV-2 from *Chlorocebus pygerythrus* is consistent with a cross-species transmission event. The lack of host-phylogeny recapitulation by the *Papio* viruses (supplementary fig. 3, Supplementary Material online) may reflect the ongoing transmission among captive baboons, which are often housed in multispecies facilities (Rogers et al. 2003) (see Supplementary Material online for a detailed discussion). In contrast, the *Macaca* viruses from each of the five represented species (*M. fascicularis, fuscata, mulatta, nemestrina,* and *silenus*) are reciprocally monophyletic and the relationships among them accurately recapitulate the host phylogeny (fig. 2B).

Molecular Clock Dating Analysis

Calibrating viral diversification events using a molecular clock in primate simplexviruses is complicated by the saturation of clock-like signal as one moves deeper into the viral phylogeny (Wertheim et al. 2014). Therefore, we calibrated a strict molecular clock using only shallow diversification events

Table 1. Inferred Substitution Rates Based on *Macaca* Calibration Priors from Three Different Published Sources.

Calibration	Fabre		Perelman		Springer	
Prior width[a]	Narrow	Wide	Narrow	Wide	Narrow	Wide
Mean[b]	7.91E-9	1.09E-8	9.11E-9	1.08E-8	9.19E-9	1.06E-8
95% HPD[b]	7.65E-9, 8.17E-9	8.55E-9, 1.33E-8	8.82E-9, 9.40E-9	9.02E-9, 1.27E-8	8.90E-9, 9.49E-9	9.29E-9, 1.20E-8

[a]Width of calibration prior on *Macaca* nodes. Narrow is 10^{-4}; wide corresponds to the published confidence interval.
[b]Substitutions/site/year.

Table 2. Inferred Time of Most Recent Common Ancestor (TMRCA) in Millions of Years Ago (Ma) for HSV-1, HSV-2, and Other Homininae Herpes Simplexviruses.

Calibration	Fabre		Perelman		Springer	
Prior width[a]	Narrow	Wide	Narrow	Wide	Narrow	Wide
Taxa[b]						
HSV-1/Great ape	11.82 (11.31–12.30)	8.71 (6.89–10.67)	10.26 (9.89–10.70)	8.73 (7.30–10.27)	10.16 (9.77–10.56)	8.84 (7.69–9.94)
HSV-2/Great ape	3.23 (3.05–3.41)	2.38 (1.85–2.90)	2.80 (2.66–2.95)	2.38 (1.98–2.81)	2.78 (2.63–2.92)	2.41 (2.10–2.75)
HSV-1	0.60 (0.55–0.65)	0.44 (0.35–0.54)	0.52 (0.48–0.56)	0.44 (0.36–0.53)	0.52 (0.48–0.55)	0.45 (0.39–0.52)
HSV-2	0.12 (0.10–0.15)	0.09 (0.07–0.12)	0.11 (0.09–0.13)	0.09 (0.07–0.11)	0.11 (0.09–0.13)	0.09 (0.07–0.11)

[a]Width of calibration prior on *Macaca* nodes. Narrow is 10^{-4}; wide corresponds to the published 95% confidence interval.
[b]Great ape viruses represent the TMRCA of ChHV, novel gorilla simplexvirus, and novel bonobo virus with either HSV-1 or HSV-2.

(<4 Ma) corresponding with host-divergence events in the clade with the closest match between the host and virus topology: McHV-1 in *Macaca*. We applied *Macaca* calibrations from three different published studies of primate divergence times (Fabre [Fabre et al. 2009]; Perelman [Perelman et al. 2011]; Springer [Springer et al. 2012]; supplementary fig. 4, Supplementary Material online), exploring the impact of both narrow and wide calibration priors (see Materials and Methods for details). We consistently inferred a substitution rate on the order of 10^{-8} substitutions/site/year (table 1), congruent with previous rate estimates from primate simplexviruses (Norberg et al. 2011; Wertheim et al. 2014). Accordingly, we inferred TMRCAs around 0.5 Ma for HSV-1 and 0.1 Ma for HSV-2 (table 2), which are broadly in agreement with previous estimates (Norberg et al. 2011; Wertheim et al. 2014; Burrel et al. 2017).

The inferred TMRCA for HSV-1 and all other Homininae simplexviruses ranged between around 7 and 12 Ma, which is consistent with the split between *Homo* and another extant Hominidae genera: *Pan*, *Gorilla*, or *Pongo* (Raaum et al. 2005) (table 2 and fig. 3). This timing suggests that HSV-1 arose via codivergence. The inferred TMRCA for HSV-2 and all other Homininae simplexviruses (excluding HSV-1) was around 2–3 Ma, far more recent than the speciation between *Homo* and *Pan* genera. This recency suggests that HSV-2 arose via cross-species transmission. We note that the ages inferred under narrow calibrations priors tended to be slightly older than ages inferred under wide priors.

The TMRCA for HSV-2 and the novel bonobo simplexvirus was bimodal due to the ambiguous phylogenetic relationship between these viruses. If these two viruses are monophyletic, as they are in around 89% of posterior trees, then we infer a median TMRCA less than 1 Ma (table 3). However, if the novel bonobo simplexvirus is more closely related to ChHV and the novel gorilla simplexvirus, as they are in around 11% of posterior samples, then HSV-2 is equally divergent from all three Homininae herpes simplexviruses.

Clock-Like Signal
To determine whether the assumption of a strict molecular clock is appropriate for a hypothesis testing framework, we repeated the molecular clock phylogenetic inference assuming a relaxed clock using an uncorrelated lognormal distribution (Drummond et al. 2006). We found no evidence for significant deviations from a strict molecular clock. Across the three calibration scenarios, the 95% highest posterior density (HPD) for the standard deviation of the lognormal distribution (a proxy for deviation from a strict clock) included or abutted against zero.

Cross-Validation Sensitivity Analysis
Our molecular clock calibrations were based on four calibrated nodes from McHV-1 infecting members of *Macaca*. We explored the internal consistency of these calibrations on the simplexvirus phylogeny using a leave-one-out cross-validation approach. The inferred TMRCAs of the great ape simplexviruses and HSV-1 or HSV-2 were robust when wide calibration priors were applied (supplementary fig. 5, Supplementary Material online). As expected, these TMRCAs were more sensitive when narrow calibration priors were applied. Excluding the calibration representing the common ancestors of 1) *M. mulatta* and *fuscata* or 2) *M. mulata*, *fuscata*, and *fascicularis* tended to produce younger TMRCAs among the Homininae simplexviruses. Excluding the narrow calibration on *M. silenus* and *nemestrina* resulted in older Homininae simplexviruses TMRCAs using the Fabre and
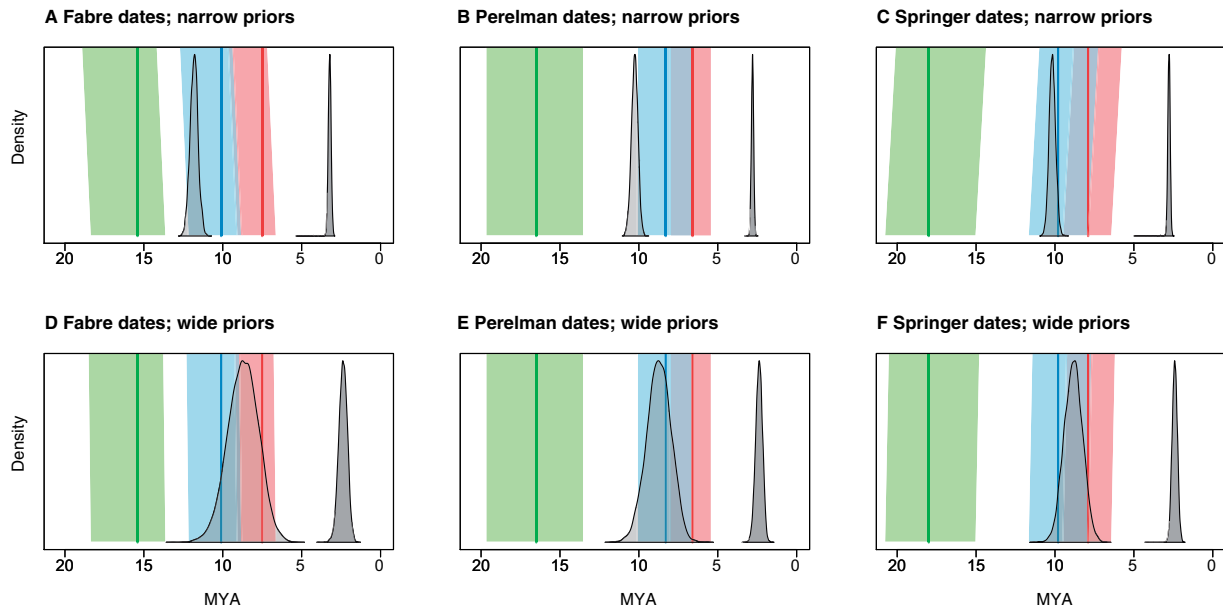
Fig. 3. Time of most recent common ancestor (TMRCA) estimates for Homininae simplexviruses (excluding humans) and HSV-1 and HSV-2. TMRCAs were inferred based on narrow calibration priors within MCHV-1 from (*A*) Fabre, (*B*) Perelman, and (*C*) Springer and from wide calibration priors within McHV-1 from (*D*) Fabre, (*E*) Perelman, and (*F*) Springer. HSV-1 TMRCA 95% highest posterior density (HPD) interval shown in light gray; HSV-2 95% HPD interval shown in dark gray. Colored solid lines represent published TMRCA estimates for *Homo* and *Pan* (red), *Gorilla* (blue), and *Pongo* (green); shaded colored blocks represent corresponding published 95% confidence intervals.

Table 3. TMRCA of HSV-2 and Novel Bonobo Herpes Simplexvirus by Different Phylogenetic Relationships in the Posterior Distribution of Tree Topologies.

Publication	Calibrations	Monophyletic			Paraphyletic		
		Posterior	TMRCA Median[a]	TMRCA 95% HPD[a]	Posterior	TMRCA Median[a]	TMRCA 95% HPD[a]
Fabre	Narrow	0.89	0.99	0.18–2.30	0.11	3.22	2.97–3.55
	Wide	0.88	0.72	0.13–1.69	0.12	2.37	1.79–2.96
Perelman	Narrow	0.89	0.86	0.18–2.03	0.11	2.80	2.65–2.95
	Wide	0.92	0.75	0.13–1.74	0.08	2.39	1.96–2.81
Springer	Narrow	0.87	0.88	0.19–2.01	0.13	2.78	2.58–2.98
	Wide	0.89	0.75	0.15–1.75	0.11	2.41	2.05–2.75

[a]TMRCAs in millions of years ago.

Perelman calibrations; whereas excluding the calibration for all *Macaca* species resulted in older Homininae simplexviruses TMRCAs only when using the Springer calibrations.

Hypothesis Testing in Generalized Stepping Stone Framework

In order to formalize our understanding of primate simplexvirus evolution, we performed a series of hypothesis tests using Bayes Factors (BF), based on marginal likelihood estimates (MLEs) obtained via a Generalized Steeping Stone (GSS) approach. The first of these tests was a direct comparison of strict versus relaxed molecular clocks. In all three calibration scenarios, we observed a BF <20 in favor of relaxing the molecular clock: Fabre (BF = 4.6), Perelman (BF = 5.0), and Springer (BF = 15). Therefore, we can confidently test hypotheses concerning the history of these herpes simplexviruses under the assumption of a strict molecular clock.

We performed a series of tests to determine whether we could reject scenarios in which the divergence of HSV-1 or HSV-2 from other simplexviruses represents a host divergence event. We added an additional constraint to the TMRCA of either HSV-1 or HSV-2 and the other Homininae simplexviruses corresponding to the estimated divergence events between *H. sapiens* and either *P. troglodytes*, *G. beringei*, or *P. pygmaeus* using a narrow calibration prior, according to the published dated primate phylogenies (Fabre et al. 2009; Perelman et al. 2011; Springer et al. 2012). As in the molecular clock comparison, we assessed the support for models using BF based on MLEs obtained via a GSS approach. These BF values represent the comparisons of models with *Macaca*-only calibrations against models with *Macaca* and an additional Hominidae calibration. In order to prevent the poor phylogenetic signal for the placement of novel bonobo simplex virus and other viral sequence fragments, these calibrations were only placed on the ancestors of the partial-genome

sequences. We included *P. pygmaeus* divergence dates in this analysis, because even though an orangutan simplexvirus has never been characterized, the lineage leading to the Homininae simplexviruses could represent *Pongo* divergence.

Using the GSS framework, we can strongly reject nearly every scenario in which the divergence of HSV-2 and a Homininae simplexvirus corresponds to a host divergence event (BF >100; fig. 4; supplementary table 1, Supplementary Material online). The evidence for rejection was only moderately strong using the Fabre dates with wide calibration constraints for a scenario in which the most recent common ancestor of HSV-2 and the Homininae simplexviruses corresponded with divergence between *H. sapiens* and *P. troglodytes* (BF = 91). Hence, it is unlikely that HSV-1 is of zoonotic origin.

The most consistently supported scenario is one in which the split between HSV-1 and the Homininae simplexviruses represents the divergence between *H. sapiens* and *G. beringei* (fig. 4; BF <20). This scenario was never rejected using wide calibration priors (fig. 4B; BF <20). Although it was weakly rejected (BF = 51) using the Perelman dates with narrow calibration priors, this scenario still had the highest MLE of any hypothesis tested using the Perelman dates with narrow calibration priors (fig. 4A).

When applying narrow calibration priors (fig. 4A), we can reject our previously hypothesized scenario (Wertheim et al. 2014) in which the most recent common ancestor of HSV-1 and Homininae simplexviruses corresponds with the divergence between *H. sapiens* and *P. troglodytes* (BF >100), except when using the Springer dates (BF = 71). Similarly, when applying the narrow calibration priors, we can strongly reject this ancestor representing the split between *H. sapiens* and *Pongo* spp. (BF >100), except when using the Fabre dates (BF = 74).

When applying wide calibration priors (fig. 4B), we cannot reject our previously hypothesized scenario in which the most recent common ancestor of HSV-1 and Homininae simplexviruses corresponds with the divergence between *H. sapiens* and *P. troglodytes* (BF <20). Similarly, we cannot reject that this ancestor corresponds with the divergence between *H. sapiens* and *Pongo* spp. using the Fabre calibrations (BF <20). There is less support for the *Pongo* codivergence scenario using the Perelman (BF = 40) and Springer (BF = 71) calibrations.

Discussion

We report the discovery of herpes simplexviruses in wild members of three extant African great ape (Homininae) species: bonobo (*P. paniscus*), mountain gorilla (*G. beringei beringei*), and western chimpanzee (*P. troglodytes verus*). Phylogenetic analysis is consistent with each Homininae species being infected with its own specific herpes simplexvirus. Although the herpes simplexviruses in Homininae are all closely related, their phylogeny is not compatible with a pattern of strict codivergence. Rather the Homininae herpes simplexviruses appear to have been repeatedly transmitted between African great ape species. Molecular clock analysis

supports a scenario in which HSV-1 arose via codivergence, with the preponderance of evidence favoring its divergence from other Homininae simplexvirus corresponding with the evolutionary divergence between *H. sapiens* and *G. beringei* lineages. HSV-2, in contrast, is likely of zoonotic origin and shares a common ancestor with a novel bonobo herpes simplexvirus in the last million years.

In our previous study into the origins of HSV-1 and HSV-2 (Wertheim et al. 2014), we constructed a hypothesis testing framework that implicitly assumed that the divergence between ChHV and either HSV-1 or HSV-2 corresponded with the divergence between *P. troglodytes* and *H. sapiens*. This framework led us to predict the existence of a novel gorilla herpes simplexvirus and that this virus would lie basal to both HSV-2 and ChHV on a phylogenetic tree. The latter half of this prediction proved to be incorrect. Rather, the more inclusive hypothesis testing framework implemented here suggests that the divergence between HSV-1 and other Homininae simplexviruses (including ChHV) more likely corresponds with the divergence between the genera *Gorilla* and *Homo*. Nonetheless, we cannot definitively reject scenarios in which this viral ancestor corresponds with the divergence between *Homo* and *Pongo* or *Homo* and *Pan* lineages.

The novel bonobo simplexvirus is most likely the closest relative to HSV-2, and these two viruses likely diverged within the past 1 My. The maximum likelihood phylogeny and the bulk of the posterior trees (≥89%) from our Bayesian inference favor a scenario in which the viruses transmitted between *H. sapiens* and *P. paniscus* sometime after *P. paniscus* diverged from *P. troglodytes* around 2.1 Ma (Stone et al. 2010; Bjork et al. 2011). However, we cannot determine the directionality of transmission based on the phylogeny alone, and it is possible that the bonobo simplexvirus arose via reverse-zoonosis from a human ancestor. Further, this phylogenetic relationship is based on a single nonsynonymous synapomorphy, so we cannot discard a scenario in which *Gorilla* and *Pan* simplexviruses are monophyletic. Therefore, the specific precursor to modern humans that was involved in the transmission event or events remains ambiguous (Wertheim et al. 2014; Houldcroft et al. 2017).

We accounted for uncertainty in divergence time estimates by separately exploring three different sets of published calibrations and by incorporating various levels of precision into these calibrations. This level of precision in our calibration priors (wide vs. narrow, see Materials and Methods) is then magnified by uncertainty in branch length and substitution rate inference inherent in our phylogenetic analysis. Nonetheless, the zoonotic origin of HSV-2, rather than HSV-1, is supported across all three sets of calibrations and both levels of precision.

We do not find evidence for the alternative scenario for the origins of the HSV-1 and HSV-2 in which the virus duplicated within an ancestral great ape species and each descendent species had two simplexviruses (as outlined and argued against in Wertheim et al. [2014]). The duplication scenario predicts that the divergence between HSV-2 and another simplexvirus corresponds with a host divergence event, which our hypothesis testing does not support. The duplication
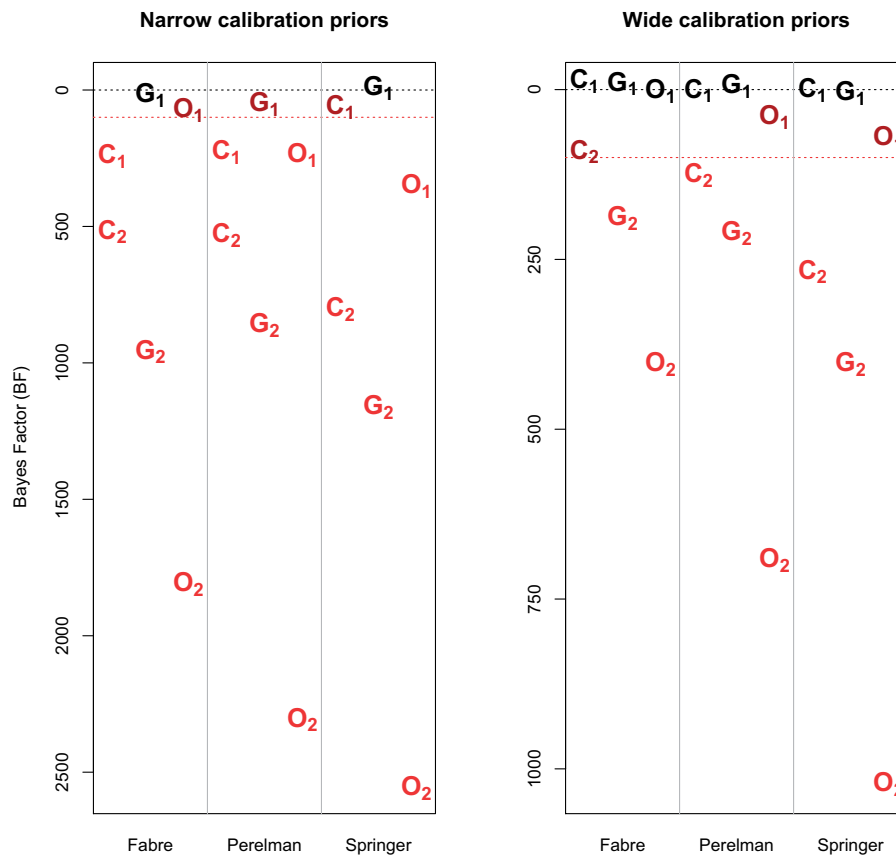
FIG. 4. Bayes factor (BF) denoting difference in marginal likelihood estimate (MLE) in the generalized stepping stone (GSS) analysis. Letters indicate the primate divergence event constrained (i.e., C, chimpanzee; G, gorilla; and O, orangutan). Subscript letters denote the human herpes simplexvirus divergence event (e.g., G_1, HSV-1/novel gorilla simplexvirus split constrained to the divergence between humans and gorillas; C_2, HSV-2/ChHV split constrained to the divergence between humans and chimpanzees). BF compares model with *Macaca*-only calibrations against models with *Macaca* and an additional Hominidae calibration. Dashed black line indicates BF=0. Dashed red line indicates BF=100. Scenarios with BF <20 are colored black, scenarios with BF ≥20 and <100 are colored dark red, and scenarios with BF >100 are colored in red.

scenario also predicts two divergent simplexviruses within all Homininae species. Although our sampling of simplexvirus diversity in Homininae is sparse, we found no indication of more than one simplexvirus infecting any African great species. If this sampling is robust, the phylogeny is consistent with a scenario in which Homininae simplexviruses have competitively excluded their predecessor after cross-species transmission. The serological similarity of these HSV-2 and the Homininae simplexviruses (Eberle 1992; Luebcke et al. 2006) may have prevented the persistence of closely related simplexviruses. Notably, we cannot exclude the possibility that some Homininae simplexviruses failed to diverge with their host species, leaving a vacant niche for a related simplexvirus to exploit (Johnson et al. 2003). Nonetheless, the evolutionary distance between HSV-1 and HSV-2 may facilitate their ability to simultaneously infect different niches in the same human host.

Estimates of deep divergence times in viruses using phylogenetics can be biased by strong purifying selection, resulting in the underestimation of branch lengths in a phylogenetic tree (Wertheim and Kosakovsky Pond 2011; Aiewsakun and Katzourakis 2016). Our current approach avoided a reliance

on molecular clock calibrations deeper in the primate phylogenetic tree. Rather, we calibrated the molecular clock using recent evidence of codivergence between five *Macaca* species and their simplexviruses. By calibrating evolutionary rates for Homininae viruses using the pattern of codivergence within the *Macaca* viruses, we were able to infer the evolutionary rate using the recent portion of the phylogeny unlikely to be biased by purifying selection. In this context, the estimated age of internal nodes deep in the phylogeny will not affect our rate estimates inferred under a strict molecular clock. If, however, the biases related to purifying selecting are present deeper in the Homininae viruses, the true TMRCA of HSV-2 and the nonhuman Homininae simplexviruses would shift even younger, favoring the scenario in which HSV-2 is the result of zoonotic transmission.

Our hypothesis testing framework used Hominidae divergence dates estimated from these same publications, rather than using speciation dates estimated from other sources. Alternate methods of dating the divergence between Hominidae species (e.g., slower pedigree-derived rates) would result in different absolute ages for the divergence events between primates (Langergraber et al. 2012; Chintalapati

and Moorjani 2020). However, our approach is robust to alternative dating schemes, because it relied on the internal consistency within a phylogeny, rather than absolute divergence times. We acknowledge that the deep divergence dates among primates provided by Perelman tend to be older than typically reported (dos Reis et al. 2018); however, the *Macaca* and Homininae calibrations employed here are medial compared with other published estimates. Given this potential for uncertainty, we note that a slight increase in the substitution rate in Homininae simplexviruses would tend to favor a *Pan* origin of HSV-2, whereas a somewhat slower substitution rate would favor a *Pongo* origin of HSV-2.

Fecal samples have been a noninvasive, productive source of knowledge about many infectious agents of primate species, more or less irrespective of the agent's suspected cellular tropism (Liu et al. 2010; Sharp and Hahn 2011; Calvignac-Spencer et al. 2012). Although the results of this first large-scale fecal screening clearly suggest noninvasive detection of viruses capable of establishing latency in neural tissues (such as simplexviruses) is very challenging, it is indeed possible. Another potential source of simplexvirus nucleic acids might be brain material from necropsies performed on naturally deceased great apes in the wild, which is however a very rare resource (Hoffmann et al. 2017). We anticipate that additional fecal and necropsy samples from gorillas, bonobos, chimpanzees, and potentially orangutans will further clarify the origin of human herpes simplexviruses.

Materials and Methods

African Great Ape Fecal Sample Screening and Genome Sequencing

In total, 1,273 fecal samples were collected at 19 sites in ten sub-Saharan countries from seven great ape species/subspecies (supplementary table 2, Supplementary Material online): *Pan troglodytes ellioti* ($n = 159$), *Pan troglodytes troglodytes* ($n = 116$), *Pan troglodytes schweinfurthii* ($n = 79$), *Pan troglodytes verus* ($n = 420$), *Pan paniscus* ($n = 135$), *Gorilla gorilla gorilla* ($n = 148$), *Gorilla beringei beringei* ($n = 216$). Authorizations of sampling were obtained from responsible local and national authorities. Except for *P. troglodytes verus* from Taï National Park, *P. paniscus* from Salonga National Park and *G. beringei beringei*, feces were obtained opportunistically from unhabituated individuals. All unhabituated *P. troglodytes* fecal samples except those from Loango National Park were collected under the umbrella of the Pan African Programme: The Cultured Chimpanzee (Kuhl et al. 2019). We did not attempt to determine the number of unhabituated individuals that were sampled. All *P. troglodytes verus* samples from Taï National Park and a fraction of *G. beringei beringei* fecal samples were obtained from pregnant chimpanzees ($n = 222$, representing 40 pregnancies by 27 individuals) and gorillas ($n = 98$, representing ten pregnancies and gorillas), respectively, under the assumption that pregnancy may favor active herpetic disease and, following, virus shedding (De Nys et al. 2014). Finally, 24 *G. beringei beringei* samples were collected from eight individuals with skin lesions suggestive of herpetic disease. DNA was extracted

using the Stool DNA Kit (Roboklon, Berlin, Germany). As part of our routine quality control, about 5% of the extracts were assessed for the presence of PCR inhibitors using an in-house qPCR assay (Calvignac-Spencer et al. 2013); no inhibition was detected.

To screen DNA extracts for simplexvirus genetic material, we used a seminested PCR approach targeting short fragments of the glycoprotein B (gB; UL27) gene: the first round oligonucleotides—HS-1f 5′-gCRggAggTggACgAgATg-3′ and HS-1r 5′-gCCAggTAgTACTgCRSCTg-3′—targeted a 225-bp fragment, within which a 160-bp fragment was targeted by the second round oligonucleotides—HS-2f-B9 5′-CSSCTCSTTCCgMTTCTC-3′ and HS-2r 5′-SAYgTgCgTSSCgTTgTA-3′. PCRs were carried out in a total volume of 25 μl and seeded with 3 μl DNA extract (first round) or 1 μl of the first round PCR product diluted 40 times (second round). Reactions contained 0.2 mM dNTP (with dUTP replacing dTTP), 4 mM MgCl$_2$, 0.2 μM of each primer, 1.25 U Platinum Taq Polymerase (Invitrogen), and 2.5 μl 10× PCR Buffer (Invitrogen, Carlsbad, CA). Cycling condition were the same for the two rounds: 95 °C 5 min, 40 cycles (95 °C 30 s, 58 °C 30 s, 72 °C 30 s), and 72 °C 10 min. In case a sample was positive with this assay, we tried to increase the gB sequence length by using two additional seminested PCR systems under the same conditions as aforementioned: 1) using first round oligonucleotides B3 (5′-TTCACCGTGGSCTGGGACTGG-3′) and HS-1r (280 bp) and second round oligonucleotides B3 and HS-2r (260 bp), 2) using first round oligonucleotides HS-1f and B10m (5′-GAGGASGTGGTCTTGATGCGYTCCACG-3′) (413 bp) and HS-2f and B10m (380 bp); and we also tried to generate amplicons for another coding sequence (UL53) using a seminested PCR system which was also run as aforementioned, using the first round oligonucleotides UL53-f1 (5′-CCSGTSACCTTCYTGTACC-3′) and UL53-r1 (5′-GCCKCTGRATCTCCTGYTCGTA-3′) (390 bp) and the second round oligonucleotides UL53-f1 and UL53-r2 (5′-ATNCCSGASAGGATGATGGA-3′) (310 bp). PCR products were cleaned up with ExoSAP-IT (Affymetrix, Santa Clara, CA) and sequenced in both directions according to Sanger's method using the BigDye Terminator kit v3.1 (Thermofischer, Waltham, MA). All chromatograms were evaluated using the software Geneious Prime (Biomatters Ltd., Auckland, New Zealand) (Kearse et al. 2012).

PCR positive extracts were fragmented using a Covaris S220 Focused-Ultrasonicator in a total volume of 130 μl (filled with low EDTA TE buffer), using settings aiming to generate approximately 400-bp fragments (intensity = 4, duty cycle = 10%, cycles per burst = 200, treatment time = 55 s, temperature = 7 °C). Fragmented DNAs were then concentrated using a MinElute PCR purification kit and eluted into 2 × 10 μl low EDTA TE buffer. DNA concentration was measured using a Qubit dsDNA High Sensitivity kit. About 1 μg DNA or all available remaining DNA extract were used for subsequent library preparation using the Accel-NGS 2S DNA library kit following the standard protocol and a sample specific unique index (Swift Biosciences, Ann Arbor, MI). Quantification was conducted using a KAPA HiFi Library

Quantification Kit (Roche, Basel, Switzerland) and libraries were then amplified using a KAPA Hot Start Library Amplification Kit (Roche) and Illumina adapter specific primers (5′-AATGATACGGCGACCACCGA-3′ and 5′-CAAGCAGAAGACGGCATACGA-3′), 45 s at 98 °C, variable number of cycles (15 s at 98 °C, 30 s at 65 °C, 45 s at 72 °C), 1 min at 72 °C. Following amplification libraries were requantified to ensure the desired amount of starting material for capture was available (>100 ng). Libraries were pooled per individual and separate hybridization capture reactions were set up. We designed 80mer RNA baits to span the RefSeq genomes of the three human alphaherpesviruses (HSV-1: NC_001806, HSV-2: NC_001798, human herpesvirus 3: NC_001348) with a 2× tiling density, which we used following the Mybaits Sequence Enrichment for Targeted Sequencing protocol (Version 2.3.1; Arbor Biosciences, Ann Arbor, MI). Following an initial round of capture, capture products were quantified and reamplified to generate 100–500 ng starting material for a second round of capture. After reamplification, the final capture products were diluted to 4 nM and sequenced on an Illumina MiSeq (v3 2x300 Chemistry; Illumina, San Diego, CA).

Raw reads were filtered (adapter removal and quality filtering) using Trimmomatic (Bolger et al. 2014) with the following settings: LEADING: 30 TRAILING: 30 SLIDINGWINDOW: 4:30 MINLEN: 30. Filtered reads were imported in Geneious, where overlapping paired-end reads were first merged using BBmerge (Bushnell et al. 2017). Merged and unmerged reads were then mapped onto the HSV-2 RefSeq genome, from which we had removed terminal inverted repeats, using the built-in mapper with default settings. The resulting map was exported and deduplicated using the SAMtools command rmdup (Li et al. 2009). The final map was imported in Geneious and used to call a consensus genome for which an unambiguous base call required that at least ten unique reads covered a position, of which 90% had to agree on the identified base. Coding sequences were identified in the consensus by similarity to coding sequences in the HSV-2 RefSeq genome, using a permissive 60% similarity threshold.

Sequence Data

Herpes simplexvirus sequence data from Catarrhine species (including Homininae and Cercopithecidae) were downloaded from GenBank. Full-length viral genomes were available from ten host species: humans, *Homo sapiens*; vervet monkeys, *Chlorocebus pygerythrus*; chimpanzees, *Pan troglodytes*; crab-eating macaques, *Macaca fascicularis*; rhesus macaques, *Macaca mulatta*; southern pig-tailed macaques, *Macaca nemestrina*; lion-tailed macaques, *Macaca silenus*; chacma baboons, *Papio ursinus*; yellow baboons, *Papio cynocephalus*; and olive baboon, *Papio anubis*. We limited the number of HSV-1 and HSV-2 sequences to representatives of major clades (Burrel et al. 2017). In addition, we included three glycoprotein B (UL27) gene fragments from McHV-1 infecting Japanese macaques, *Macaca fuscata*.

Many of the McHV-1 and PaHV-2 sequences were obtained from captive animals that had been housed with related *Macaca* and *Papio* species. Therefore, we investigated the provenance of each virus to determine if 1) the host species could not be reliably determined due to cohousing of multiple species belonging to the same genus or 2) the virus detected in one species in captivity likely infected a different wild species. See Supplementary Material online for a detailed description and supplementary table 3, Supplementary Material online, for accession numbers and references.

Alignment

We identified genes in the novel gorilla simplexvirus that contained at least 500 nonambiguous nucleotides and had homologous genes in both Homininae and Cercopithecidae simplexviruses. Homologous coding regions in these related viruses were defined according to their GenBank annotations. Sequences for each gene were aligned separately as translated amino acids in Muscle v3.8 in Aliview v1.23 (Edgar 2004; Larsson 2014). We then added the partial glycoprotein B (UL27) viral sequences from *Macaca fuscata* (267 bp for two sequences and 1,944 bp for the other), *P. troglodytes verus* (124 bp), *P. paniscus* (124 bp), and *G. beringei beringei* (124 bp for two sequences and 365 bp for the other); glycoprotein K (UL53) sequence from two *G. beringei beringei* were also included (272 bp). Genes with known recombinant regions within alphaherpes simplexvirus (i.e., UL15, UL29, UL30, and UL39; Burrel et al. 2017) were excluded and the remaining 29 genes were concatenated into a single partial-genome alignment.

Recombination Detection

Recombination is pervasive in human herpes simplexviruses, including between HSV-1 and HSV-2 (Lamers et al. 2015; Burrel et al. 2017) and among PaHV-2 (Tyler and Severini 2006). The genomic regions analyzed here were restricted to regions with high-sequencing coverage in the novel gorilla simplexvirus genome and known recombinant regions in HSV-2 (UL15, UL29, UL30, and UL39) were excised from the alignment (Burrel et al. 2017). We screened the 74 taxon alignment for recombination using RDP4 (Martin et al. 2015) and found no robust evidence for recombination between these viral lineages in the alignment. However, we cannot exclude the possibility of recombination between the novel gorilla simplexvirus and other Homininae simplexviruses in other genomic regions.

The final alignment included 74 taxa and, after removing recombinant regions and genes with <500-bp coverage in the novel gorilla simplexvirus, was 43,797 bp in length.

Maximum Likelihood Phylogenetic Analysis

We inferred a simplexvirus maximum likelihood phylogenetic tree using IQTree v1.6.3 under a GTR+Γ_4 substitution model (Nguyen et al. 2015). Node support was determined using 1,000 ultrafast bootstrap replicates (Hoang et al. 2018). Tree was visualized using FigTree v1.4.3.

Selection Analysis

A test for diversifying selection was performed using BUSTED with synonymous site variation on the Datamonkey server (Murrell et al. 2015; Weaver et al. 2018; Wisotsky et al. 2020).

Bayesian Molecular Clock Analysis

Bayesian phylogenetic inference was conducted using BEAST v1.10.4 (Suchard et al. 2018). The bias in TMRCA estimation deep in viral phylogenies due to purifying selection has been previously demonstrated to strongly affect branch length estimation separating the major catarrhine simplexvirus lineages (Wertheim et al. 2014); however, the shallower divergences within this phylogeny were not affected. Therefore, we calibrated the molecular clock using divergence events in the McHV-1 clade. This calibration is based on the assumption of codivergence within the McHV-1 viruses and their corresponding five *Macaca* host species, because the McHV-1 phylogenetic relationships perfectly recapitulated those of their hosts. We identified three published studies on primate divergence times that included estimates for all four *Macaca* divergence events represented in the simplexvirus phylogeny: Fabre, Perelman, and Springer (supplementary fig. 4, Supplementary Material online). These studies inferred *Macaca* TMRCAs using different molecular clock approaches based on different sources of external calibrations. Further, each of the studies also provided estimates for TMRCAs within Hominidae, which are necessary for subsequent hypothesis testing.

We explored two different sets of priors based on the estimated node ages from the literature: narrow (normal distribution with standard deviation of 1×10^{-4}) and wide (normal distribution encompassing reported confidence intervals from the literature). We opted to use the reported node ages as prior distributions, rather than incorporate the fossil data itself, because of the sparsity of known primate simplexviruses relative to the number of corresponding primate fossils. Each of these six scenarios was run in duplicate for a Markov chain Monte Carlo (MCMC) length of 10 million generations, sampling ever 10,000 generations. We used a GTR+Γ_4 substitution model assuming empirical base frequencies in a birth–death framework employing strict or relaxed (uncorrelated lognormal distribution) molecular clock. Convergence (effective sample sizes ≥200) and mixing were assessed in Tracer v1.7 (Rambaut et al. 2018), after removing the first 10% generations as burn-in. A maximum clade credibility (MCC) tree displaying median node heights was generated using TreeAnnotator. Subsampling of the posterior TMRCA distributions of HSV-2 and the novel bonobo simplexvirus based on monophyly and paraphyly was achieved using the ETE3 package (Huerta-Cepas et al. 2016).

Cross-Validation Analysis

A leave-one-out cross-validation sensitivity analysis was performed in order to assess the robustness of the four *Macaca* calibrations. Sensitivity was assessed on the TMRCAs between HSV-1/HSV-2 and the Homininae herpes viruses. As in the primary MCMC analysis, cross-validation was performed using both narrow and wide calibration priors.

Model Testing Using GSS Marginal Likelihood Estimator

We performed a series of hypothesis tests to determine whether the TMRCAs of Homininae simplexviruses and HSV-1 or HSV-2 corresponded to the TMRCA of *H. sapiens* and *Pan*, *Gorilla*, or *Pongo*. Even though we do not have evidence for the existence of an orangutan (*Pongo*) herpes simplexvirus, it has been hypothesized that the split between HSV and Homininae simplexviruses may actually represent this deeper codivergence event. To incorporate this testing into the MCMC framework, for each of the three published calibration schemes (Fabre, Perleman, and Springer), we included additional calibration priors at the ancestor of either 1) HSV-1, HSV-2, ChHV, and the novel gorilla simplexvirus or 2) HSV-2, ChHV, and the novel gorilla simplexvirus. These priors were applied as normal distributions with narrow standard deviations (10^{-4}), as wider calibrations were too accommodating. Taxa with only partial-gene sequences were excluded from the calibration prior specifications but maintained in the phylogenetic analysis.

We also compared the marginal likelihood estimates between the strict and relaxed molecular clock analyses, as described above.

MLEs are intuitive estimations of model fit that are founded on probability (Oaks et al. 2019). To compare the likelihood of each calibration set, we selected a GSS approach (Fan et al. 2011; Baele et al. 2016), a streamlined development of path- (Lartillot and Philippe 2006) and stepping stone- (Xie et al. 2011) sampling which accommodates phylogenetic uncertainty (Baele et al. 2012). We applied a burn-in of one million from the original MCMC, an MLE chain length of 250,000, and 75 path steps. Optimal chain length was determined through exploratory analysis on a single set of calibrations varying chain length (10,000–1,000,000) and path steps (50–150). Working priors from the MCMC were applied under the destination of the MLE. GSS analyses with joint, prior, or likelihood estimated sample size <200 or duplicate runs whose MLEs differed >15 BF, suggesting poor convergence, were repeated. Due to the variability of MLEs from duplicate runs (Baele et al. 2016), we used a significance threshold of 100 BF for establishing strong evidence for rejecting a model.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Data Availability

Raw reads and assembled sequences are available at the European Nucleotide Archive under study accession PRJEB39611. Alignments, trees, and XML input files can be accessed on Data Dryad at https://datadryad.org/stash/data-set/doi:10.6076/D13S3X.

References

Aiewsakun P, Katzourakis A. 2016. Time-dependent rate phenomenon in viruses. *J Virol.* 90(16):7184–7195.
Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol.* 29(9):2157–2167.
Baele G, Lemey P, Suchard MA. 2016. Genealogical working distributions for Bayesian model testing with phylogenetic uncertainty. *Syst Biol.* 65(2):250–264.
Bjork A, Liu W, Wertheim JO, Hahn BH, Worobey M. 2011. Evolutionary history of chimpanzees inferred from complete mitochondrial genomes. *Mol Biol Evol.* 28(1):615–623.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
Burrel S, Boutolleau D, Ryu D, Agut H, Merkel K, Leendertz FH, Calvignac-Spencer S. 2017. Ancient recombination events between human herpes simplex viruses. *Mol Biol Evol.* 34(7):1713–1721.
Bushnell B, Rood J, Singer E. 2017. BBMerge – accurate paired shotgun read merging via overlap. *PLOS One* 12(10):e0185056.
Calvignac-Spencer S, Leendertz SA, Gillespie TR, Leendertz FH. 2012. Wild great apes as sentinels and sources of infectious disease. *Clin Microbiol Infect.* 18(6):521–527.
Calvignac-Spencer S, Merkel K, Kutzner N, Kuhl H, Boesch C, Kappeler PM, Metzger S, Schubert G, Leendertz FH. 2013. Carrion fly-derived

DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Mol Ecol.* 22(4):915–924.
Chintalapati M, Moorjani P. 2020. Evolution of the mutation rate across primates. *Curr Opin Genet Dev.* 62:58–64.
Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC, Pellett PE, Roizman B, Studdert MJ, Thiry E. 2009. The order Herpesvirales. *Arch Virol.* 154(1):171–177.
De Nys HM, Calvignac-Spencer S, Boesch C, Dorny P, Wittig RM, Mundry R, Leendertz FH. 2014. Malaria parasite detection increases during pregnancy in wild chimpanzees. *Malar J.* 13:413.
de Vienne DM. 2019. Tanglegrams are misleading for visual evaluation of tree congruence. *Mol Biol Evol.* 36(1):174–176.
dos Reis M, Gunnell GF, Barba-Montoya J, Wilkins A, Yang Z, Yoder AD. 2018. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. *Syst Biol.* 67(4):594–615.
Drummond AJ, Ho SY, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4(5):e88.
Dudas G, Carvalho LM, Rambaut A, Bedford T. 2018. MERS-CoV spillover at the camel-human interface. *eLife* 7:e31257.
Duggan AT, Perdomo MF, Piombino-Mascali D, Marciniak S, Poinar D, Emery MV, Buchmann JP, Duchene S, Jankauskas R, Humphreys M, et al. 2016. 17(th) century variola virus reveals the recent history of smallpox. *Curr Biol.* 26(24):3407–3412.
Dux A, Lequime S, Patrono LV, Vrancken B, Boral S, Gogarten JF, Hilbig A, Horst D, Merkel K, Prepoint B, et al. 2020. Measles virus and rinderpest virus divergence dated to the sixth century BCE. *Science* 368(6497):1367–1370.
Eberle R. 1992. Evidence for an alpha-herpesvirus indigenous to mountain gorillas. *J Med Primatol.* 21(5):246–251.
Eberle R, Black D. 1993. Sequence analysis of herpes simplex virus gB gene homologs of two platyrrhine monkey alpha-herpesviruses. *Arch Virol.* 129(1–4):167–182.
Eberle R, Hilliard JK. 1989. Serological evidence for variation in the incidence of herpesvirus infections in different species of apes. *J Clin Microbiol.* 27(6):1357–1366.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792–1797.
Fabre PH, Rodrigues A, Douzery EJ. 2009. Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear DNA. *Mol Phylogenet Evol.* 53(3):808–825.
Fan Y, Wu R, Chen MH, Kuo L, Lewis PO. 2011. Choosing among partition models in Bayesian phylogenetics. *Mol Biol Evol.* 28(1):523–532.
Grose C. 2012. Pangaea and the out-of-Africa model of Varicella-Zoster virus evolution and phylogeography. *J Virol.* 86(18):9558–9565.
Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35(2):518–522.
Hoffmann C, Zimmermann F, Biek R, Kuehl H, Nowak K, Mundry R, Agbor A, Angedakin S, Arandjelovic M, Blankenburg A, et al. 2017. Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. *Nature* 548(7665):82–86.
Hon CC, Lam TY, Shi ZL, Drummond AJ, Yip CW, Zeng F, Lam PY, Leung FC. 2008. Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. *J Virol.* 82(4):1819–1826.
Hoppe E, Pauly M, Gillespie TR, Akoua-Koffi C, Hohmann G, Fruth B, Karhemere S, Madinda NF, Mugisha L, Muyembe JJ, et al. 2015. Multiple cross-species transmission events of human adenoviruses (HAdV) during hominine evolution. *Mol Biol Evol.* 32(8):2072–2084.
Houldcroft CJ, Ramond JB, Rifkin RF, Underdown SJ. 2017. Migrating microbes: what pathogens can tell us about population movements and human evolution. *Ann Hum Biol.* 44(5):397–407.
Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol Biol Evol.* 33(6):1635–1638.
Johnson KP, Adams RJ, Page RD, Clayton DH. 2003. When do parasites fail to speciate in response to host speciation? *Syst Biol.* 52(1):37–47.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.

Kuhl HS, Boesch C, Kulik L, Haas F, Arandjelovic M, Dieguez P, Bocksberger G, McElreath MB, Agbor A, Angedakin S, et al. 2019. Human impact erodes chimpanzee behavioral diversity. *Science* 363(6434):1453–1455.

Lamers SL, Newman RM, Laeyendecker O, Tobian AA, Colgrove RC, Ray SC, Koelle DM, Cohen J, Knipe DM, Quinn TC. 2015. Global diversity within and between human herpesvirus 1 and 2 glycoproteins. *J Virol.* 89(16):8206–8218.

Langergraber KE, Prufer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, et al. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci U S A.* 109(39):15716–15721.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22):3276–3278.

Lartillot N, Philippe H. 2006. Computing Bayes factors using thermodynamic integration. *Syst Biol.* 55(2):195–207.

Leroy EM, Kumulungui B, Pourrut X, Rouquet P, Hassanin A, Yaba P, Delicat A, Paweska JT, Gonzalez JP, Swanepoel R. 2005. Fruit bats as reservoirs of Ebola virus. *Nature* 438(7068):575–576.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, Ndjango JB, Sanz CM, Morgan DB, Locatelli S, et al. 2010. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. *Nature* 467(7314):420–425.

Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 395(10224):565–574.

Luebcke E, Dubovi E, Black D, Ohsawa K, Eberle R. 2006. Isolation and characterization of a chimpanzee alphaherpesvirus. *J Gen Virol.* 87(Pt 1):11–19.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1(1):vev003.

McGeoch DJ, Cook S, Dolan A, Jamieson FE, Telford EA. 1995. Molecular phylogeny and evolutionary timescale for the family of mammalian herpesviruses. *J Mol Biol.* 247(3):443–458.

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, et al. 2015. Gene-wide identification of episodic selection. *Mol Biol Evol.* 32(5):1365–1371.

Nelson MI, Worobey M. 2018. Origins of the 1918 pandemic: revisiting the swine "Mixing Vessel" hypothesis. *Am J Epidemiol.* 187(12):2498–2502.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.

Norberg P, Tyler S, Severini A, Whitley R, Liljeqvist JA, Bergstrom T. 2011. A genome-wide comparative evolutionary analysis of herpes simplex virus type 1 and varicella zoster virus. *PLOS One* 6(7):e22527.

Oaks JR, Cobb KA, Minin VN, Leache AD. 2019. Marginal likelihoods in phylogenetics: a review of methods and applications. *Syst Biol.* 68:681–697.

Perelman P, Johnson WE, Roos C, Seuanez HN, Horvath JE, Moreira MA, Kessing B, Pontius J, Roelke M, Rumpler Y, et al. 2011. A molecular phylogeny of living primates. *PLoS Genet.* 7(3):e1001342.

Raaum RL, Sterner KN, Noviello CM, Stewart CB, Disotell TR. 2005. Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J Hum Evol.* 48(3):237–257.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* 67(5):901–904.

Rogers KM, Ealey KA, Ritchey JW, Black DH, Eberle R. 2003. Pathogenicity of different baboon herpesvirus papio 2 isolates is characterized by either extreme neurovirulence or complete apathogenicity. *J Virol.* 77(20):10731–10739.

Severini A, Tyler SD, Peters GA, Black D, Eberle R. 2013. Genome sequence of a chimpanzee herpesvirus and its relation to other primate alphaherpesviruses. *Arch Virol.* 158(8):1825–1828.

Sharp PM, Hahn BH. 2011. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med.* 1(1):a006841.

Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghwani J, Bhatt S, et al. 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 459(7250):1122–1125.

Smith JS, Robinson NJ. 2002. Age-specific prevalence of infection with herpes simplex virus types 2 and 1: a global review. *J Infect Dis.* 186(Suppl 1):S3–28.

Springer MS, Meredith RW, Gatesy J, Emerling CA, Park J, Rabosky DL, Stadler T, Steiner C, Ryder OA, Janecka JE, et al. 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS One* 7(11):e49521.

Stone AC, Battistuzzi FU, Kubatko LS, Perry GH Jr, Trudeau E, Lin H, Kumar S. 2010. More reliable estimates of divergence times in Pan using complete mtDNA sequences and accounting for population structure. *Philos Trans R Soc Lond B Biol Sci.* 365(1556):3277–3288.

Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4(1):vey016.

Tyler SD, Severini A. 2006. The complete genome sequence of herpesvirus papio 2 (Cercopithecine herpesvirus 16) shows evidence of recombination events among various progenitor herpesviruses. *J Virol.* 80(3):1214–1221.

Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, Sahl JW, Enk J, Birdsell DN, Kuch M, Lumibao C, et al. 2014. *Yersinia pestis* and the plague of Justinian 541-543 AD: a genomic analysis. *Lancet Infect Dis.* 14(4):319–326.

Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol.* 35(3):773–777.

Weiss RA. 2007. Lessons from naked apes and their infections. *J Med Primatol.* 36(4–5):172–179.

Wertheim JO, Kosakovsky Pond SL. 2011. Purifying selection can obscure the ancient age of viral lineages. *Mol Biol Evol.* 28(12):3355–3365.

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(9):2356–2364.

Wisotsky SR, Kosakovsky Pond SL, Shank SD, Muse SV. 2020. Synonymous site-to-site substitution rate variation dramatically inflates false positive rates of selection analyses: ignore at your own peril. *Mol Biol Evol.* 37(8):2430–2439.

Xie W, Lewis PO, Fan Y, Kuo L, Chen MH. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol.* 60(2):150–160.