Intrahost speciations and host switches shaped the evolution of herpesviruses

ANDERSON F. BRITO[1*]; JOHN W. PINNEY[1*]
1. Department of Life Sciences, Imperial College London, South Kensington Campus. London, SW7 2AZ. United Kingdom

* Corresponding authors: Anderson F. Brito - andersonfbrito@gmail.com
John W. Pinney - j.pinney@imperial.ac.uk

ABSTRACT

Cospeciation has been suggested to be the main force driving the evolution of herpesviruses, with viral species co-diverging with their hosts along more than 400 million years of evolutionary history. Recent studies, however, have been challenging this assumption, showing that other co-phylogenetic events, such as intrahost speciations and host switches play a central role on their evolution. Most of these studies, however, were performed with undated phylogenies, which may underestimate or overestimate the frequency of certain events. In this study we performed co-phylogenetic analyses using time-calibrated trees of herpesviruses and their hosts. This approach allowed us to (i) infer co-phylogenetic events over time, and (ii) integrate crucial information about continental drift and host biogeography to better understand virus-host evolution. We observed that cospeciations were in fact relatively rare events, taking place mostly after the Late Cretaceous (~100 Millions of years ago). Host switches were particularly common among alphaherpesviruses, where at least 10 transfers were detected. Among beta- and gammaherpesviruses, transfers were less frequent, with intrahost speciations followed by losses playing more prominent roles, especially from the Early Jurassic to the Early Cretaceous, when those viral lineages underwent several intrahost speciations. Our study reinforces the understanding that cospeciations are uncommon events in herpesvirus

evolution. More than topological incongruences, mismatches in divergence times were the main disagreements between host and viral phylogenies. In most cases, host switches could not explain such disparities, highlighting the important role of losses and intrahost speciations in the evolution of herpesviruses.

Keywords: tree reconciliation; herpesvirus; phylogenetics; host transfers; virus-host evolution

INTRODUCTION

Herpesviridae is a diverse family of large double-stranded DNA viruses subdivided in three subfamilies – *Alpha*-, *Beta*-, and *Gammaherpesvirinae* –, which infect different groups of vertebrates, including birds, mammals and reptiles (Davison et al. 2009). The evolutionary history of herpesviruses (HVs) dates from the Early Devonian, around at least 400 Millions of years ago (Mya) (McGeoch et al. 2006). Up to recent years, HVs were considered to be species-specific, and to have evolved alongside their hosts mainly by cospeciation (McGeoch et al. 1995; Davison 2002; Jackson 2005; McGeoch et al. 2006), an event that implies concurrent and interdependent splits of host and viral lineages over time (de Vienne et al. 2013). As more information about the divergence timing of host and viral species became available, the predominance of cospeciation as the main mechanism of HV evolution started to be challenged, especially due to mismatches between the divergence times of hosts and viruses, which evoked alternative hypothesis to explain HV evolution, such as host switches (transfers) (Davison 2002; Escalera-Zamudio et al. 2016).

Transfers take place when viruses succeed in infecting a new host still unexplored by their ancestors (de Vienne et al. 2013). Recent studies have been suggesting that host switches are probably more frequent than previously thought (Escalera-Zamudio et al. 2016; Geoghegan et al. 2017), but detecting host switches can be difficult, as extinctions

of viruses transmitted to new hosts may occur frequently (Geoghegan et al. 2017). Duplication of viral lineages (intrahost speciation) is another evolutionary process playing an important role on host-pathogen evolution (de Vienne et al. 2013). By means of intrahost speciations, multiple species of viruses can explore a single host species, especially when new lineages occupy distinct biological niches (tissues) within the hosts (Davison 2002). Finally, another event playing a key role in host-virus evolution is the loss of viral lineages, events that in a co-phylogenetic context can mean: (i) symbiotic extinction (Lovisolo et al. 2003); (ii) sorting events (Johnson et al. 2003); or even (iii) rare/undiscovered species, as a result of undersampling (Page and Charleston 1998).

The aforementioned mechanisms can be inferred using co-phylogenetic analyses, such as tree reconciliations, which help us understand the relationship between hosts and parasites over time (Page and Charleston 1998). To do that, such methods identify differences and similarities between the topologies of host and parasite trees, where congruence may indicate points of co-divergence (cospeciation), while incongruences my imply host switches or intrahost speciations followed by losses (de Vienne et al. 2013). Topological congruence, however, not always is caused by cospeciation events: similar topologies of host and parasite phylogenies can happen by chance, as a result of repeated host switches (de Vienne et al. 2013). To better understand the intricate evolutionary processes of HVs, co-phylogenetic analyses can be applied to elucidate the pathways taken by viruses while their hosts and the environment evolve. By integrating time-calibrated phylogenies and historical biogeographical information, in this study we present detailed scenarios unravelling the evolution of herpesviruses with their hosts, providing answers to the following questions: (*i*) What are the main co-phylogenetic events shaping HV evolution?; (*ii*) How have herpesviruses achieved their broad host range?; (*iii*) Have the three HV subfamilies evolved following similar strategies?

RESULTS

Phylogenetic analysis

Herpesvirus evolutionary history is older than that of their extant hosts (McGeoch et al. 2006). The MRCA (Most Recent Common Ancestors) of all hosts date back between 352 and 307 Mya, while the MRCA of all HV subfamilies existed at least between 416 and 373 Mya. As evidenced by the tanglegram in Figure 1, despite the apparent congruence of some clades in both phylogenies, topological disagreements are common for most virus-host pairs, as highlighted by the overcrossing connections between the trees.

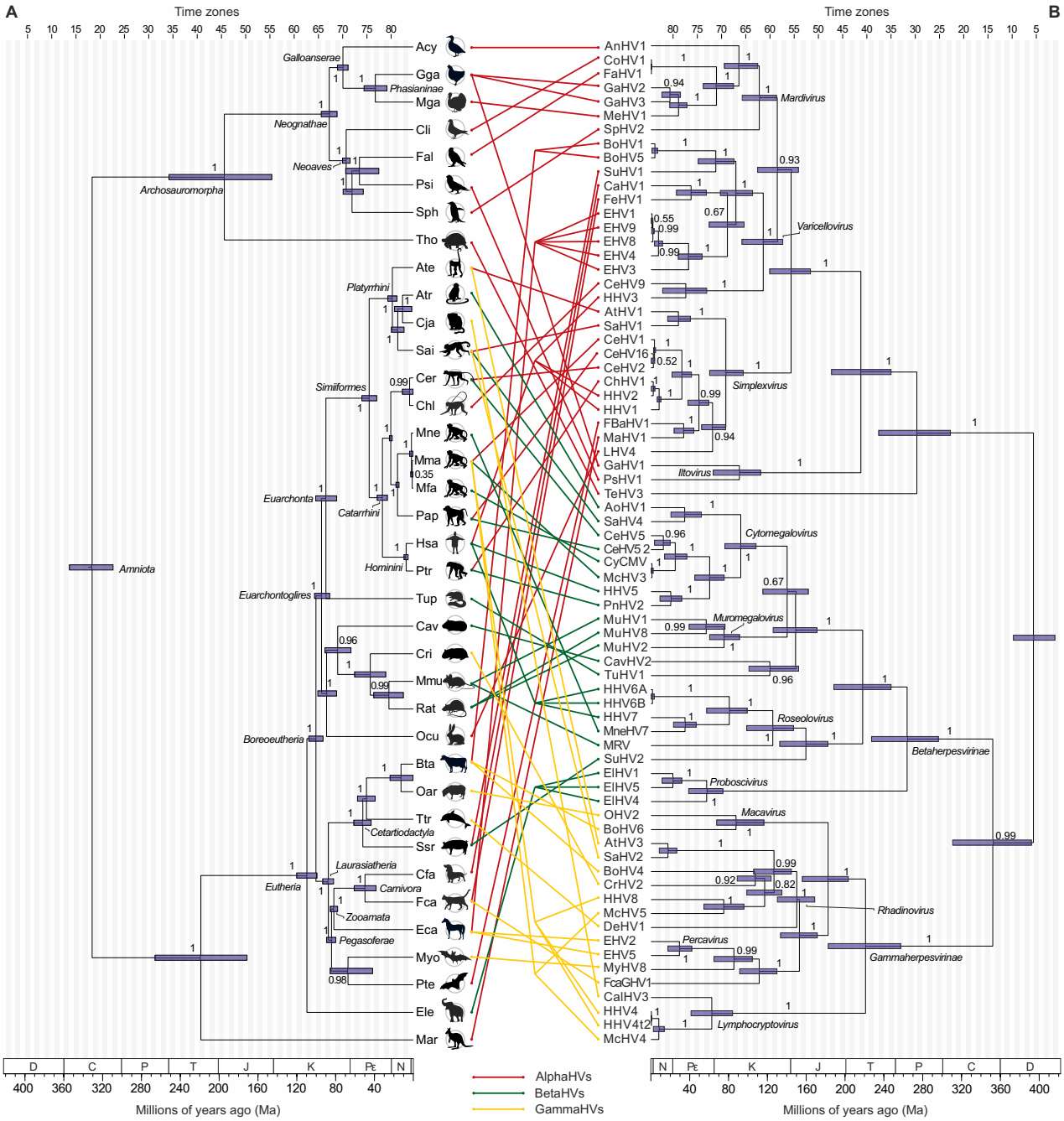


Figure 1. Host-virus tanglegram. Viruses (right) are connected to their hosts (left) with lines whose colours denote the three herpesviral subfamilies: *Alpha-* (red); *Beta-* (green) and *Gammaherpesvirinae* (yellow). Node height HPD intervals of hosts and viruses are shown as bars, labels are provided for some taxonomic groups, and all nodes are assigned with posterior probabilities. All internal nodes in the host tree were time-calibrated with priors. In the viral tree, priors were assigned to only two nodes (the root node and HHV1-HHV2 MRCA), while the others had their date ranges estimated from sequence data. Both trees are divided in time zones of 5 Myr, as shown by the scale at the top, and some node HPD intervals span more than one time zone. The geologic time scale is set according to (Gradstein et al. 2012), where D = Devonian period; C = Carboniferous; P = Permian; T = Triassic; J = Jurassic; K = Cretaceous; Pε = Paleogene; N = Neogene; and * = Quaternary period. Host acronyms are defined as follows: Acy = *Anser cygnoides*; Ate = *Ateles sp.*; Atr = *Aotus trivirgatus*; Bta = *Bos taurus*; Cav = *Cavia porcellus*; Cer = *Cercopithecus aethiops*; Cfa = *Canis lupus familiaris*; Chl = *Erythrocebus patas*; Cja =

Callithrix jacchus; Cli = *Columba livia*; Cri = *Cricetidae*; Eca = *Equus caballus*; Ele = *Elephas maximus*; Fal = *Falco mexicanus*; Fca = *Felis catus*; Gga = *Gallus gallus*; Hsa = *Homo sapiens*; Mar = *Macropodidae*; Mfa = *Macaca fascicularis*; Mga = *Meleagris gallopavo*; Mma = *Macaca mulatta*; Mmu = *Mus musculus*; Mne = *Macaca nemestrina*; Myo = *Myotis velifer*; Oar = *Ovis aries*; Ocu = *Oryctolagus cuniculus*; Pap = *Papio sp*.; Psi = *Amazona oratrix*; Pte = *Pteropus sp*.; Ptr = *Pan troglodytes*; Rat = *Rattus sp*.; Sai = *Saimiri sp*.; Sph = *Spheniscus sp*.; Ssr = *Sus scrofa*; Tho = *Testudo horsfieldii*; Ttr = *Tursiops truncatus*; and Tup = *Tupaiidae*. For more details on host and viral taxonomy, accession numbers, and TaxID can be found in S1 Table.

Cost regimes

The co-phylogenetic analysis of 72 herpesviral species and their respective 37 hosts has revealed different possible scenarios to explain their evolution, according to distinct event costs. We tested 256 cost regimes, which favoured or penalized events differently (see supplementary tables S2-S4). Since the association of internal nodes in viral and host trees was constrained by their time zone ranges, the total number of possible solutions was limited, with many cost regimes resulting in similar solutions, which differed mostly in terms of the time zones where the inferred events probably took place. As shown in Figure 2, except for AlphaHVs (Figure 2A), the number of possible solutions (i.e. the number of events inferred for each event type, in distinct cost regimes) was very restricted in reconciliations involving BetaHVs and GammaHV, with both showing two main scenarios (Figure 2B and C): one with high numbers of Intrahost Speciations (IS) and Losses (LO), resulting in less likely solutions; and another one where such events were inferred at much lower numbers, mostly matching the median values (grey horizontal bars). For AlphaHVs, such fluctuation between high and low numbers of IS and LO was also observed, but more continuously distributed at both extremes.

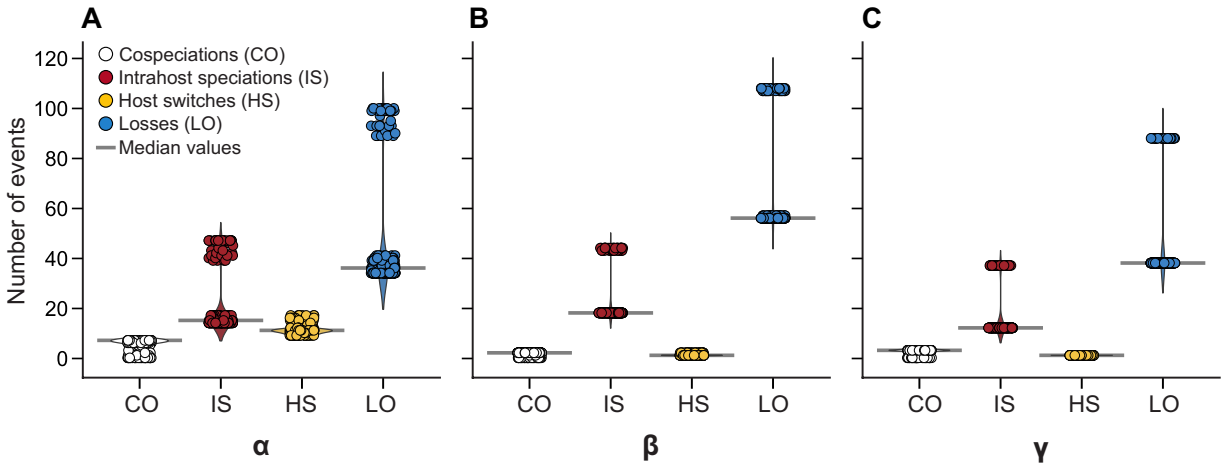


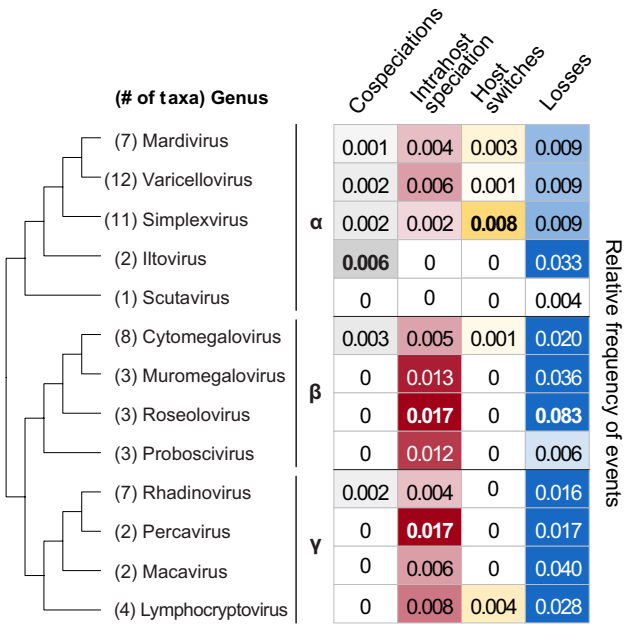
Figure 2. Co-phylogenetic events inferred in reconciliations between virus and host trees. A total of 256 distinct cost regimes were explored, resulting in distinct, but relatively similar scenarios explaining the disagreements between viral and host evolutionary histories. The main differences among the inferred scenarios concerned the number of Intrahost Speciations (IS, red) and Losses (LO, blue), two closely associated events. The panels show the number of inferred events under distinct cost regimes, in virus-host tree reconciliations involving A) AlphaHVs, B) BetaHVs and C) GammaHvs. Based on the median number of events for each event type (grey bars), an optimal cost regime was selected.

For each subfamily, a single cost regime was selected according to the following order of criteria, it should: (1) reconstruct the median number of events for each event type (as shown by grey bars on Figure 2); (2) ensure the highest possible support values for reconstructed events, and; (3) output the lowest possible overall cost. As a result, an optimal cost regime with the following relative costs was selected to explain the topological disagreements between viral and host tree: Cospeciation (CO) = 0; Intrahost speciation (IS) = 1; Host Switch (HS) = 2; and Loss (LO) = 0. Since the reconciliations between HVs from different subfamilies were analysed separately, it allowed us to examine the predominance of each co-phylogenetic event across different HV genera (Figure 2), as shown in the following sections.

Cospeciations over the last 100 millions of years

Cospeciations were only found after the Late Cretaceous (~100 Mya), and have shown to be a rare event among HVs (Figure 2 and 3). For alphaherpesviruses, seven

152 cospeciation events were reconstructed (Table 1), the oldest one assigned to the Late
153 Cretaceous (~92-81 Mya), involving ancestors of *Varicellovirus* infecting *Laurasiatheria*
154 mammals (see Figure 4). This event was followed by at least other two cospeciations, one
155 at the split of ancestors of canines, felines and equines (*Zooamata*); and another at the
156 split of *Carnivora* ancestors. Among mardiviruses, ancestors of GaHV1 and PsHV1 (genus
157 *Iltovirus*) co-diverged with bird ancestors (*Neognathae*) around 78-94 Mya, similar to what
158 occurred for viruses infecting ancestors of chickens and turkeys (*Phasianinae*) around 36-
159 26 Mya. Finally, among simplexviruses infecting *Catarrhini* (Old World Monkeys and
160 Apes), two cospeciations were found, one taking place most likely at the Oligocene (~36-
161 26 Mya); and the second at the Miocene (~9-5 Mya), involving viruses infecting ancestors
162 of Humans and Chimpanzees (*Hominini*).



163

164 **Figure 3. Relative frequency of co-phylogenetic events in distinct herpesviral genus,**
165 **considering the optimal cost regime.** Values are normalized by the total number of taxa in each
166 genus, and their respective times to the MRCA. As highlighted, losses were the most frequent co-
167 phylogenetic events, followed by intrahost speciations.

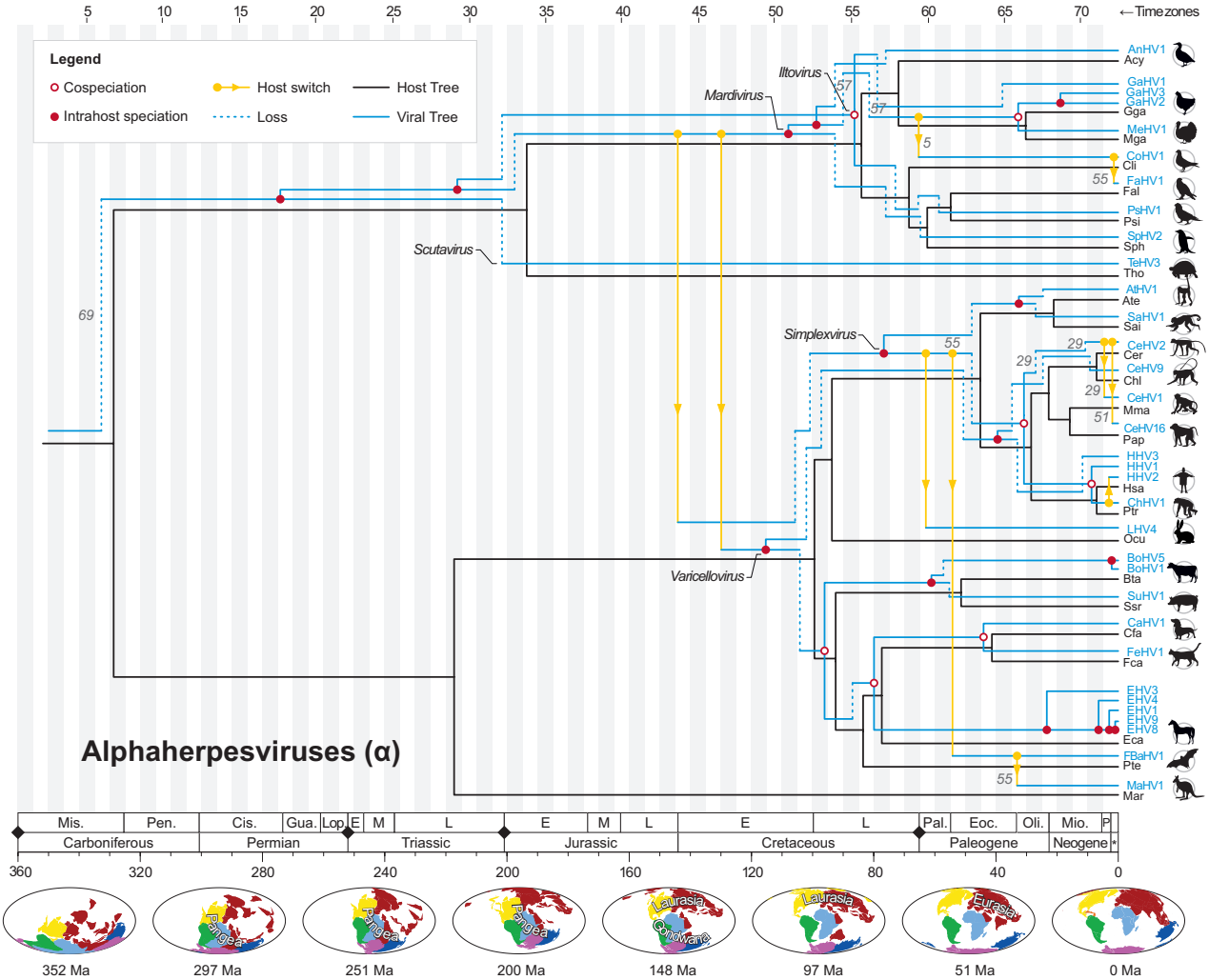


Figure 4. Tree reconciliation of Alphaherpesviruses and their hosts. In this representation, the host tree is shown in black, and the viral tree is shown in blue, vertically twisted, keeping its original topology and assigned time zones as shown in Figure 1. Co-phylogenetic events with support values (%) below 100 are shown. Along the time scale, black diamonds denote major events of mass extinction, as described in (Raup 1993). The maps at the bottom show changes of landmasses (continental drift) over time, and were retrieved from PBDB (Peters and McClennen 2015). Cis. = Cisuralian; E = Early; Eoc. = Eocene; Gua. = Guadalupian; L = Late; Lop. = Lopingian; M = Middle; Mio. = Miocene; Mis. = Mississippian; Oli. = Oligocene; P = Pliocene; Pal. = Paleocene; Pen. = Pennsylvanian; and * = Quaternary.

Table 1. Overall statistics of co-phylogenetic events inferred under the optimal cost regime used for virus-host tree reconciliations. The number of inferred events per HV subfamily (α, β and γ) match the median values in Figure 1. Cospeciation = (CO); Intrahost speciation = (IS); Host Switch = (HS); and Loss = (LO).

Subfamily	Number of inferred events				Overall cost
	CO	IS	HS	LO	
α	7	15	10	36	35
β	2	18	1	56	20
γ	3	12	1	38	14

In Betaherpesviruses, only two cospeciations were reported, both along the lineage of cytomegaloviruses infecting ancestors of *Cercopithecine* (macaques and baboons) (Figure 5): the first one at the Paleogene-Neogene boundary (~24-21 Mya); and the second one around 0.8 Mya, the most recent cospeciation found in this study, involving ancestors of CyCMV and McHV3, at the split between *Macaca fascicularis* (Mfa) and *M. mulatta* (Mma).

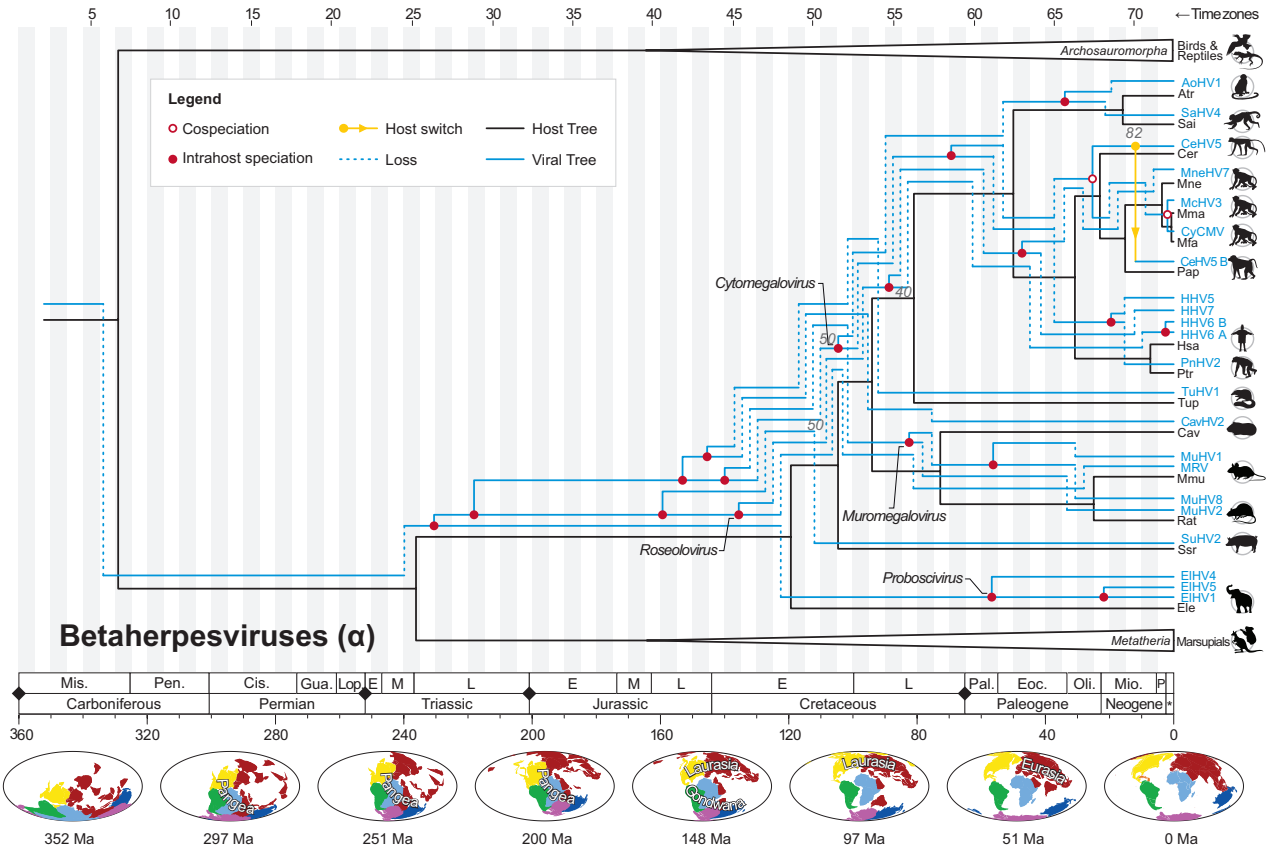


Figure 5. Tree reconciliation of Betaherpesviruses and their hosts.

Among Gammaherpesviruses, three cospeciations were found in different periods (Figure 6). The earliest event involved *Macavirus* ancestors, and took place at the split between *Laurasiatheria* and *Euarchontoglires* (Early Cretaceous, ~99 Mya). Following this event, HVs infecting ancestors of bats, equines and felines (*Pegasoferae*) co-diverged with their hosts at the Late Cretaceous (~88-79 Mya); and finally, at the Paleogene-Neogene boundary (~25-16 Mya), *Rhadinovirus* ancestors co-diverged with New World Monkeys (*Platyrrhini*).

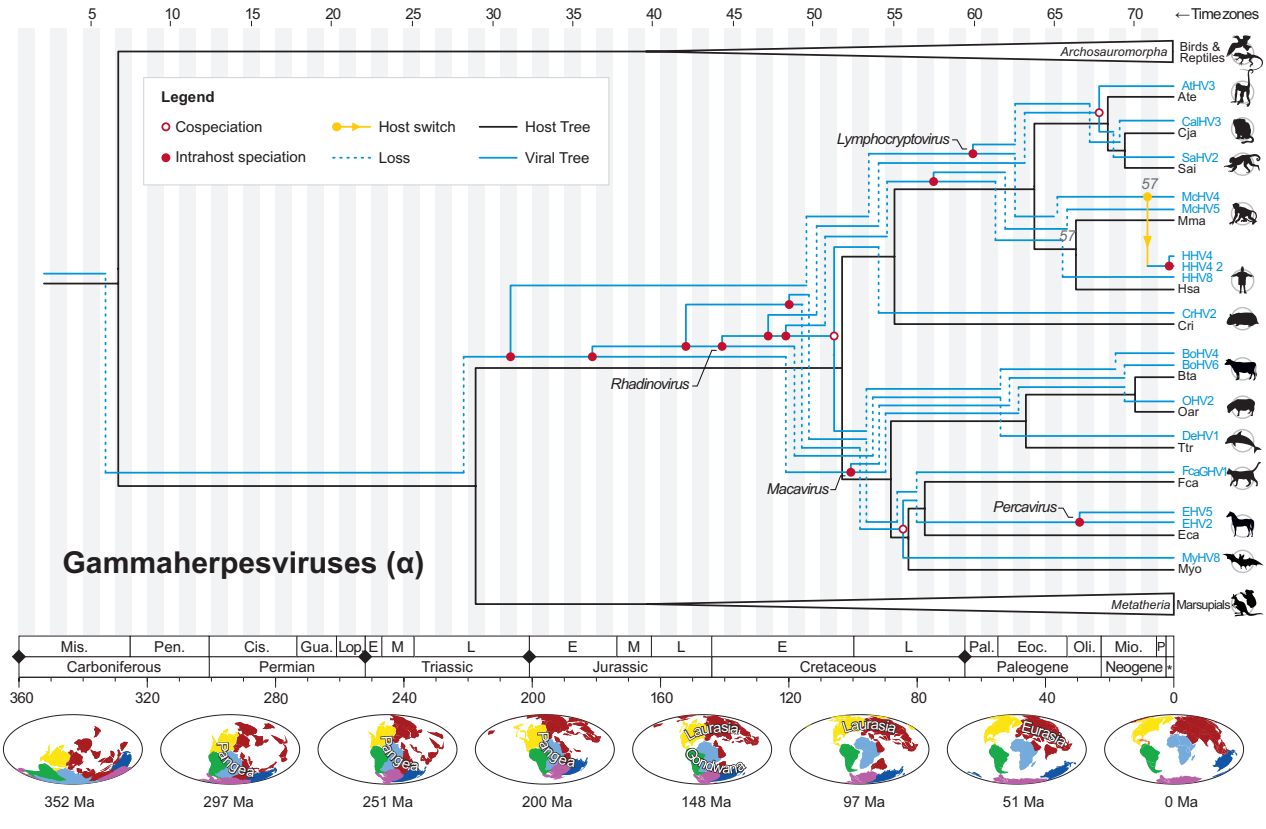


Figure 6. Tree reconciliation of Gammaherpesviruses and their hosts.

203 Intrahost speciations: duplications of viral lineages

204 Intrahost speciations (also known as duplications) occur when a parasite diverges
205 and both lineages remain within the same host species, and represent alternative modes
206 of evolution that explain mismatches between divergence times of viruses and hosts (de
207 Vienne et al. 2013). Intrahost speciation was the second most frequent co-phylogenetic
208 event in the evolution of HVs (Figure 2 and 3). Along the evolution of AlphaHVs, such
209 events were particularly common after the Early Cretaceous. During the Permian and
210 Triassic periods, when AlphaHVs infected ancestors of birds and reptiles
211 (*Archosauromorpha*), two important intrahost speciations gave rise to ancestors of the
212 main genera of *Alphaherpesvirinae* (Figure 4). Later in their evolution, along the
213 Paleogene and Neogene, these viruses underwent a process of diversification that gave
214 rise to multiple species sharing common hosts, as observed for Gallid (GaHV2, 3); Equid
215 (EHV1, 3, 4, 8, 9); and Bovine alphaherpesviruses (BHV1, 5). Differently from these recent
216 duplications, older duplication events were always followed by losses, which may explain,
217 for example, the apparent lack of scutaviruses infecting birds, or mardiviruses and
218 iltoviruses infecting reptiles. The evolutionary histories of BetaHVs and GammaHVs are
219 also largely characterized by intrahost speciations (Figure 2B and 2C). As a consequence
220 of early intrahost speciations, ancestors of BetaHV and GammaHV co-existed infecting
221 early placental mammals from the Late Triassic to the Early Cretaceous, period when such
222 viral lineages experienced extensive diversification (Figure 5 and Figure 6).

223

224 Losses: extinctions, sorting events and undiscovered herpesviruses

225 Among the four co-phylogenetic events, losses were the most common (Figure 2 and
226 3). Remarkably, considering all HV subfamilies, more than half of the losses (nearly 54%)
227 were assigned to the Cretaceous period (~145-66 Mya). As previously mentioned, losses

228 are usually preceded by intrahost speciations, and highlight host clades that lack viruses
229 from certain lineages. At this point it is essential to emphasize that in the context of host-
230 parasite tree reconciliations, losses can be interpreted in at least three distinct ways: (1) as
231 lineage sorting events ('missing the boat'), when a parasite fails to disperse to one of the
232 new host species after their speciation (Johnson et al. 2003); (2) as undiscovered or rare
233 parasites (low sampling) (Page and Charleston 1998); or (3) as genuine events of parasite
234 extinctions (Lovisolo et al. 2003). In the latter scenario, if extinctions explain the absence
235 of viruses infecting certain host clades, it is important to consider that points of losses in
236 reconciliations do not reflect the exact time of the extinctions, but rather highlight a point
237 from which such events could have happened at any subsequent time.

238 Throughout the evolution of herpesviruses, in Figure 4, Figure 5, and Figure 6, viral
239 losses are depicted as dashed lines pointing towards the opposite direction the losses are
240 assigned to. In the evolution of AlphaHVs, for example, the oldest loss dates from the
241 Carboniferous period, before the split between sauropsids (bird/reptile ancestors) and
242 synapsids (mammal ancestors) (Figure 4). This loss was inferred due to the potential lack
243 of a basal group of AlphaHVs infecting synapsids in that time. Most losses of AlphaHVs
244 were assigned to periods after the Early Cretaceous, along the earliest events of
245 diversification of modern birds and mammals. Except for *Scutavirus*, a genus here
246 represented by a single species (TeHV3), all genera of *Alphaherpesvirinae* show extensive
247 losses between 110 and 70 Mya. Proportionally, losses were the most common events in
248 the evolution of the genus *Iltovirus* (Figure 3), an old clade of HVs infecting distantly
249 related avian families: *Psittacidae* and *Phasianidae*. Another context where losses played
250 a central role was during the evolution of varicello- and simplexviruses infecting
251 *Euarchontoglires* (a group that includes rabbits and primates). The evolution of these

252 viruses alongside their hosts could only be explained by means of multiple losses,
253 especially from ~90 Mya onwards, where at least 12 losses were observed.

254 Along the evolution of Beta- and Gammaherpesviruses losses were predominant.
255 Among the BetaHV genera, *Roseolovirus* has shown the highest level of losses (Figure 3).
256 Ancestors of these viruses originated most likely in the Early Cretaceous (~146-119 Mya),
257 and across their evolution they were either lost in most host groups or may still exist as
258 rare/undiscovered viruses. In GammaHVs, the genus *Macavirus* showed the second
259 highest relative frequency of losses. Viruses in this taxonomic group are known to infect
260 *Cetartiodactyla* hosts, mostly ruminants and swine (Ehlers et al. 2008). Since the MRCAs
261 of these viruses and their *Laurasiatheria* hosts existed around 87 Mya (Late Cretaceous),
262 in case they were not extinct, macaviruses infecting *Pegasoferae* (felines, canines,
263 equines and bats) may still exist in nature.

264

265 Host switches

266 Host switches (transfers) take place when parasites succeed on infecting new hosts
267 not yet explored by their ancestors (de Vienne et al. 2013). In all scenarios (cost regimes)
268 investigated, host switches were invoked to explain herpesvirus-host evolution. Based on
269 the optimal cost regime adopted in our analyses (Table 1), at least 10 host switches were
270 reported for AlphaHVs. The oldest host switch took place when HVs from bird ancestors
271 (sauropsids) got transferred to placental mammals in the Early Cretaceous (~144 Mya),
272 before the split between *Mardivirus* and *Varicellovirus*. Later (~130 Mya), a second host
273 switch took place, giving rise to ancestors of all varicelloviruses, which currently are found
274 only in a wide range of mammalian hosts (Davison 2010) (Figure 4). Interestingly, other
275 than these two early transfers, all the remaining host switches of AlphaHVs were assigned
276 to periods after the Cretaceous-Paleogene boundary. In the *Mardivirus* lineage, viruses

now infecting falconids and columbids (*Neoaves*) have likely originated after two host transfers: one around the Paleocene (~67 Mya) involving viruses from ancestors of modern chicken and turkeys (*Phasianinae*); and a second one in more recent times, when viruses infecting pigeons got transferred to falcons in the upper Pleistocene (~150 thousands of years ago), the most recent transfer identified in this study.

Most notably, host switches were particularly common among simplexviruses, where more than 2/3 of the transfers in AlphaHVs were observed (Figure 3). In this genus, two early transfers from primates gave rise to viruses now infecting Rabbits, Bats and Marsupials: ancestors of *Leporid alphaherpesvirus* (LHV4) switched from primates to rabbits around 76-52 Mya; and later, primate HVs switched to bats in the Eocene (~54-41 Mya), with descendants of these viruses later being transferred to marsupials in the Oligocene (~33 Mya), giving rise to the present-day *Macropodid herpesvirus* (MaHV1). The remaining transfers involved the interchange of simplexviruses among Old World Monkeys and Apes (*Catarrhini*) in the Pliocene. Ancestors of HVs now infecting monkeys of the genus *Macaca* (Rhesus, infected by CeHV1), *Papio* (Baboons, infected by CeHV16), and *Cercopithecus* (Guenons, infected by CeHV2) were likely transferred among these monkey species between 3.3 and 2.1 Mya. The exact origin of their ancestors, and the polarity of the host switches were not possible to be determined only by phylogenetic analysis, as multiple possible scenarios have shown the same overall costs (see Figure 4). Finally, an important *Simplexvirus* transfer took place between 3.6 and 0.1 Mya, when viruses infecting Chimpanzees (genus *Pan*) were transferred to humans, giving rise to HHV2, as already reported in a previous study (Wertheim et al. 2014).

In Beta- and GammaHVs, host switches were less prominent, with only one event being assigned to each of these groups, both involving primate hosts. Among BetaHVs, the species here named CeHV5-B (originally known as isolate OCOM4-52 (Blewett et al.

302 2015)) originated in the Miocene (~12 Mya) from viruses transferred from Guenon
303 ancestors (genus *Cercopithecus*). Finally, for GammaHVs our results provide a possible
304 explanation for the origins of HHV4 (Epstein-Barr Virus). Among lymphocryptoviruses,
305 McHV4, a *Macacine gammaherpesvirus* infecting Rhesus macaques (*M. mulatta*), is the
306 non-human HV more closely related to HHV4 (Figure 6). Ancestors of these viruses
307 infected either Apes (early hominids) or Old World Monkeys (early cercopithecids) around
308 13.3 - 2.1 Mya, when these HVs switched hosts. For this particular event, the directionality
309 of the host transfer (hominids ↔ cercopithecids) was not possible to be determined
310 because, given the tree topologies and node heights, both directions of transfer imply the
311 same overall cost. Despite that, other than host switch, no other co-phylogenetic event can
312 explain the existence of these viruses in macaques and humans.
313
314 **DISCUSSION**
315
316 To understand the presence of herpesviruses infecting distinct host groups, co-
317 phylogenetic analyses with time-calibrated trees can provide consistent explanation for the
318 dynamics of HVs evolution alongside their hosts. By using temporal data in tree
319 reconciliations, we could also incorporate information about geological and biogeographic
320 events that shaped life on Earth. A previous study placed the origin of herpesviruses in the
321 Devonian (~400 Mya) (McGeoch and Gatherer 2005), a time reference used as one of the
322 priors to calibrate the herpesviral phylogeny in this study. Based on comparative structural
323 studies, it has been proposed that herpesviruses share a common ancestor with tailed
324 phages (Baker et al. 2005). This hypothesis may suggest an older origin for Herpesviridae,
325 however, the precise time range is not known, as these viruses do not produce identifiable
326 fossil evidences. In this way, to assign divergence times to herpesvirus trees, we made

327 use of molecular clock estimates from previous theoretical studies (McGeoch and
328 Gatherer 2005; Wertheim et al. 2014). Despite the intrinsic uncertainties and limitations of
329 this method, we consider such estimates to be the most objective way so far available for
330 calibrating large DNA virus phylogenies. For this study, we calibrated the viral tree at only
331 two nodes (the deepest node, and a shallow node), letting the remaining intermediate
332 nodes free to have their dates estimated from sequence data. By setting constrains of
333 monophyly for well characterized viral and host taxonomic groups, we inferred robust
334 phylogenies with estimated ranges of divergence times, which allowed us to reconstruct, in
335 a temporal scale, the events underlying host-viral evolution since ancient periods.

336 For a long time, cospeciation has been considered to be the main evolutionary
337 mechanism driving the evolution of herpesviruses and their hosts (McGeoch et al. 1995;
338 Davison 2002; Jackson 2005; McGeoch et al. 2006). However, most previous studies
339 performed reconciliations using trees with few taxa, and without time scale, thus ignoring
340 mismatches between divergence times of host and viral ancestors. For cospeciations to be
341 accurately inferred, pathogen and host trees must not only show topological congruence,
342 but also correspondent divergence times (de Vienne et al. 2013). Thus, the absence of
343 temporal scale at internal nodes may overestimate the occurrence of certain events as a
344 result of chronologically inconsistent node pairings. In the present study we performed
345 reconciliation analyses using time-calibrated trees, which allowed us to observe that, due
346 to temporal incompatibilities, cospeciations were in fact rare events along the herpesviral
347 evolution, with other events playing more predominant roles.

348 After extensive phylogenetic characterization of herpesviruses, topological
349 disagreements between host and viral trees became evident, and since cospeciations
350 alone were no longer enough to explain the evolution of HVs, host transfers were initially
351 presumed to account for such incongruences (Ehlers et al. 2008; Escalera-Zamudio et al.

352 2016). Contrasting with such assumption, our results revealed that while transfers played a
353 key role on *Alphaherpesvirinae* evolution, they were uncommon among *Beta*- and
354 *Gammaherpesvirinae*. For these two subfamilies, intrahost speciations followed by losses
355 were abundant (Figure 2B), and may provide better explanations for the topological
356 disagreements between their phylogenies and those of their hosts.

357 Viral losses do not necessarily mean extinctions, as they may also indicate
358 undersampling or undiscovered viruses. Since the highest levels of losses were observed
359 for herpesviral genera with few sampled taxa (Figure 3), a hypothesis of undiscovered
360 viruses appears to be plausible, as the relative frequency of losses were to some extent
361 influenced by the number of sampled taxa. If extinctions are invoked to explain these
362 losses, some events linked to the evolution of the hosts can be ascribed as potential
363 causes of the elimination of viral clades, one of them being host extinction. Although each
364 mass extinction could have wiped out from 76% to 96% of the ancient species, most
365 extinctions along the Phanerozoic Eon were the result of minor events taking place in-
366 between mass extinctions (Raup 1993). Since the average duration of species is
367 estimated to be four Myr, with genera lasting for around 28 Myr (Raup 1993), symbiont
368 extinction may explain some of the losses inferred in the present study.

369 Apart from host populations being wiped out, leading to viral elimination, cataclysmic
370 events can also cause drastic decreases in host populations (Hesse and Buckling 2016),
371 which may promote bottleneck effects directly affecting viral adaptation to hosts, and
372 leading to viral extinctions. It occurs especially because bottlenecks can increase the
373 likelihood of fixation of certain host alleles, including those granting host resistance to
374 pathogens (Hesse and Buckling 2016). Among bird species, for example, a strong
375 bottleneck impacted avian populations after the 5th mass extinction (Feduccia 2003),
376 decreasing by nearly half the net diversification rates of birds along the Paleocene and

377 Eocene (~66-45 Mya) (Claramunt and Cracraft 2015). Such decline in host diversity may
378 possibly explain some of the losses inferred in clades of avian HVs, especially those
379 belonging to *Iltovirus* (Figure 3 and Figure 4).

380 Since direct or indirect contact are required for viral spread into new hosts, the
381 geological movement of landmasses can split or merge host populations, in this way
382 allowing or preventing certain host switches (Lovisolo et al. 2003). For a better
383 understanding of ancestral host transfers, it is important to consider the biogeography
384 (spatial distribution) of ancestral hosts, and the geological history of the Earth. As shown at
385 the bottom of Figure 4, Figure 5 and Figure 6, alongside the evolution of hosts and their
386 associated viruses, the planet underwent drastic changes. All hosts included in this study
387 are tetrapods, a group of organisms that diverged from aquatic vertebrates between 385
388 and 375 Mya (George and Blieck 2011). Given that the most probable tMRCA of
389 herpesviruses dates from 416 and 373 Mya, ancestors of such viruses probably infected
390 marine organisms, and along their terrestrialization, their viruses diverged into ancestors of
391 the known Alpha-, Beta-, and GammaHVs. Being the first to diverge, AlphaHVs were
392 probably lost on synapsids (proto-mammals) in the Carboniferous, reappearing as
393 mammalian HVs much later, when avian AlphaHVs got transferred to mammals in at least
394 two independent events (Figure 4). Such transfers, which gave rise to *Simplexvirus* and
395 *Varicellovirus*, were already suggested in previous studies to explain the dispersal of HVs
396 in mammalian hosts (McGeoch and Gatherer 2005; McGeoch et al. 2006). Our previous
397 study has shown that HVs gained, duplicated and lost specific protein domains along their
398 evolutionary history (Brito and Pinney 2020). Before and after colonizing their new hosts,
399 simplexviruses acquired, duplicated and lost domains encoded mainly in envelope and
400 modulatory proteins, while varicelloviruses evolved by altering their repertoire of envelope
401 and auxiliary protein domains.

402 Although parasites are more likely to jump between closely related hosts, as they
403 share similar ecological, physiological and chemical characteristics (de Vienne et al. 2013;
404 Geoghegan et al. 2017), host switches can also occur over great phylogenetic distances,
405 as hosts from closely related clades can independently acquire or lose immunogenetic
406 traits (epitopes, domains or whole genes), which may change their levels of susceptibility
407 to pathogens (Longdon et al. 2014). As host populations could have been not only
408 physically but also genetically closer in early times than they are in current times, the
409 origins of mammalian HVs by means of host transfers from avian ancestors could be a
410 reasonable explanation.

411 In our analysis, except for the two transfers discussed above, all host switches were
412 found in more recent times, especially after the Cretaceous/Paleogene boundary (~66
413 Mya). This pattern was already expected, as pointed out in a previous study, which found
414 recent host switches as a predominant evolutionary patter in most DNA and RNA viruses
415 (Geoghegan et al. 2017). Our results revealed that host switches were more prominent
416 among simplexviruses. A remarkable example is the transfer involving FBaHV1 and
417 MaHV1, which are closely related to primate HVs, but infect distantly related hosts (bats of
418 the genus *Pteropus* and marsupials, respectively). As ancestors of *Pteropus sp.*
419 (Megabats) inhabited Eurasia alongside primate ancestors in the Paleocene and Eocene
420 (~59-38 Mya) (Springer et al. 2011; Springer et al. 2012), transfers of HVs from primates
421 to bats were probably more likely to happen in that period. Finally, the subsequent host
422 switch from megabats to macropods (Kangaroo and Wallaby ancestors) can be mainly
423 explained by their current and ancestral distributions in Australia, and by the dispersal
424 capabilities of bats using powered flight (Springer et al. 2011). As our previous study
425 revealed, both FBaHV1 and MaHV1 underwent major genomic reshaping after being
426 transferred from primates and diverge into two viral species: FBaHV1 gained and

427 duplicated a series of envelope and modulatory protein domains, while MaHV1 lost at
428 least eight domains, mainly present in envelope proteins (Brito and Pinney 2020). Given
429 the defensive and offensive strategies of evolution adopted by pathogens and their hosts
430 in molecular arms races (Daugherty and Malik 2012), herpesviruses likely succeeded at
431 host switching by evading and/or neutralizing host immune factors by acquiring and losing
432 elements from their domain repertoires (Brito and Pinney 2017; Brito and Pinney 2020).

433 Host switches involving primate HVs were observed in all herpesvirus subfamilies,
434 most of them taking place after ~10 Mya, along the Miocene, Pliocene and Quaternary.
435 Wertheim and colleagues (Wertheim et al. 2014) performed an in-depth analysis of the
436 evolution of Human simplexviruses (HHV1 and HHV2) and suggested that HHV2 could
437 have originated from host switches of HVs from Chimpanzees around 1.6 Mya. Our
438 findings have confirmed this transfer, positioning it slightly earlier in time, around 2 Mya
439 (95% HPD = 3.6 - 0.1 Mya), after the divergence of ChHV1 and HHV2. The same study
440 has also pointed out another possible host switch involving ancestors of CeHV2. Our
441 results revealed that simplexviruses infecting Old World Monkeys, such as CeHV1,
442 CeHV2, and CeHV16 (Figure 1), underwent host switches after 4.3 Mya. Remarkably, all
443 species of the genera *Cercopithecus* and *Papio* (except *P. hamadryas*) are found only in
444 Africa, while those of the *Macaca* genus are mostly found in Asia (except *M. sylvanus*)
445 (Springer et al. 2012). Ancestors of *Macaca* dispersed from Africa to Asia in the Pliocene,
446 between 5.4 and 2.3 Mya (Springer et al. 2012), a period that matches the tMRCA of
447 simplexviruses infecting Old World Monkeys. Thereby, our findings suggest that transfers
448 of HVs from *Cercopithecus* to *Macaca* took place before the migration of the latter to Asia,
449 which explains the presence of CeHV1 infecting *M. Mulatta*. Finally, between 3.5 and 0.07
450 Mya, transfers from *Cercopithecus* to ancestors of *Papio* took place in Africa, the place of

origin and current area of distribution of most species from both genera (Springer et al. 2012).

It is known that herpesviral host switches occurred frequently in the past, especially among closely related organisms (Thiry et al. 2006; Tischer and Osterrieder 2010). As mutation rate and effective population size (*Ne*) affect the likelihood of adaptation of a pathogen in a new host (Longdon et al. 2014), the extinction of viruses transmitted to new host species may occur frequently (Geoghegan et al. 2017). As a result, most host switches cannot be easily detected, and the number of transfers inferred in our co-phylogenetic analysis is probably underestimated.

In conclusion, in this study we used time-calibrated virus-host phylogenies to perform tree reconciliations. By means of this approach we were not only able to detect topological disagreements between viral and host tree topologies, but more importantly, we reveal major chronological mismatches on divergence times of animal species and their herpesviruses, showing the relevance and fundamental contribution of temporal data for reconciling phylogenies. Our dated reconciliations highlighted the central roles of intrahost speciations in the evolution of herpesviruses, events that in most cases were followed by losses, which were predominant along the Cretaceous period. It was not possible to clarify what such losses represent in the evolution of herpesviruses, as they can represent lineage sorting events, undiscovered viruses, or even episodes of viral extinctions. As more viral samples are incorporated in tree reconciliations, the real nature of such losses will be revealed. Host switches were particularly frequent among alphaherpesviruses, especially those of the genus *Simplexvirus*. Viruses from this genus and other genera deserve further studies to uncover whether the gain or loss of specific genetic traits (domains and genes) may have favoured the colonization of new hosts or tissues (Brito and Pinney 2020). Finally, contrasting what was

475 previously thought, cospeciations between herpesviruses and their hosts are rare events,
476 mainly observed among alphaherpesviruses.
477
478 **MATERIALS AND METHODS**
479
480 Host/virus species and phylogenetic analysis
481 A total of 72 herpesviruses (subfamilies *Alpha-, Beta-, and Gammaherpesvirinae*)
482 and their 37 host species (mammals, birds and reptiles) were included in this study (see
483 S1 Table). Only viral species with whole genomes available on NCBI Viral Genomes
484 Resource (Brister et al. 2015) were considered in the analysis. To infer both host and viral
485 species phylogeny, sequences of host nuclear genes and conserved viral proteins were
486 retrieved from NCBI. Sequence of genes encoding BDNF, CNR1, EDG1, RAG1, and RHO
487 were used for inferring the host tree, and the viral tree was generated using sequences of
488 orthologs of the HHV1 genes UL15, UL27 and UL30 (McGeoch et al. 2000). For host
489 genes, when species-specific sequences were not available, sequences from related
490 species from the same taxonomic group were used. Each gene set was aligned using
491 MAFFT (Katoh and Standley 2013), and models of nucleotide and amino acid substitution
492 were determined using jModelTest (Posada 2008) and ProtTest (Darriba et al. 2011),
493 respectively. The multiple sequence alignments were used as partitions in *BEAST to infer
494 maximum clade credibility (MCC) species trees using the Markov Chain Monte Carlo
495 (MCMC) Bayesian approach implemented in BEAST v2.4.5 (Bouckaert et al. 2014). The
496 analyses were performed using relaxed (uncorrelated lognormal) molecular clock, with the
497 Yule model as a coalescent prior. The host tree was run for 500 million generations, and
498 the viral one for 35 million generations.

499 The host tree had their node ages time-calibrated (in Millions of years, Myr) using as
500 priors host divergence dates obtained from (Hedges and Kumar 2009; dos Reis et al.
501 2012; Claramunt and Cracraft 2015). With that, all internal nodes in the host phylogeny
502 were assigned with lower and upper bounds for divergence times, which match the
503 divergence dates provided at timetree.org (Kumar et al. 2017). Since fossil records or
504 other ancestral evidences are not available for herpesviruses, viral divergence times were
505 calibrated using theoretical estimates derived from molecular clock analyses performed by
506 (McGeoch and Gatherer 2005; Wertheim et al. 2014). Following the same strategy used
507 by (Brito and Pinney 2020), in the herpesvirus phylogeny, only the deepest node (root)
508 and a shallow internal node representing the divergence between HHV1-HHV2 were
509 assigned with time calibration priors, while the remaining nodes were left free to have had
510 their divergence times inferred from sequence data. Constraints of monophyly were
511 applied for viral clades based on the current taxonomic classification provided by ICTV
512 (King et al. 2018), and host clades were constrained according to two recent and
513 comprehensive studies (dos Reis et al. 2012; Claramunt and Cracraft 2015). Such
514 constraints allowed the reconstruction of robust phylogenies, which fully agree with the
515 current taxonomy of viral and host groups under study.
516
517 Tree reconciliation
518 Tree reconciliations were performed separately for each HV subfamily using the
519 program Jane 4 (Conow et al. 2010). Such analyses were performed to find low cost
520 associations between host and viral trees, while keeping their original topologies. Taking
521 advantage of their node height highest posterior density (HPD) intervals, all trees were
522 converted into a Jane timed tree format using a Python script available at GitHub (see
523 Data Availability). At this step, the continuous time scales of the phylogenies were

discretized into bins of 5 Myrs (“time zones”) shared by viral and host trees. In this way, the internal nodes of host and viral trees could be assigned to specific time frames, ensuring that only nodes belonging to the same time zone could be associated for inferring potential host switches and cospeciation events, thus avoiding chronological inconsistencies. The algorithm implemented in Jane allows internal nodes to be assigned to more the one time zone, and requires that all zones should be populated with at least one host node. To meet such requirement, an outgroup clade containing artificial taxa was added to the original host tree, ensuring that their internal nodes could span time zones not originally covered by the original host nodes. A similar outgroup was added to the viral tree, allowing the pairing of artificial virus-host pairs.

Since it is not possible to ascertain the actual relative costs of events of cospeciations, intrahost speciation, host switches or losses in the evolution of HVs, we reconciled trees under multiple combinations of relative costs, in which those events were weighted with cost values varying from 0 to 3. For each HV subfamily, a total of 256 co-phylogenetic cost regimes were explored, generating several solutions, which differed in terms of overall cost and number of inferred events (see supplementary tables S2-S4). In this solution space, the median number of inferred events for each event type was calculated, and an optimal cost regime was selected based on its ability to: (i) produce the exact or similar reconciliation scenario with total number of inferred events equivalent to the median values, (ii) in which events have the highest possible supporting values, (iii) with the lowest overall cost. Following these criteria, the selected cost regime had the following relative costs of events: cospeciations = 0; intrahost speciation = 1; host switches = 2; and losses = 0 (see supplementary tables S2-S4).

ACKNOWLEDGEMENTS

AFB is funded by *Ciência sem Fronteiras*, a scholarship programme managed by the Brazilian federal government (CAPES, Ministry of Education, Grant number: 11911-13-1). JWP is supported by a University Research Fellowship from the Royal Society. The authors thank the Imperial College London Open Access Fund for the financial support.

DATA AVAILABILITY

All genome used mentioned in this study have their accession numbers listed as Supplementary Data. All data and codes generated in this study are deposited in the following repository on GitHub: https://github.com/andersonbrito/openData/tree/master/brito_2020_reconciliation.

SUPPLEMENTARY DATA

Supplementary Materials are available online. Supplementary Table S1 shows the list of genomes used in this study. Supplementary Tables S2-4 lists all cost regimes used for inferring possible scenarios explaining the evolution of Alpha-, Beta, and Gammaherpesviruses.

COMPETING INTERESTS

The authors declare that they have no competing interests.

REFERENCES

Baker ML, Jiang W, Rixon FJ, Chiu W. 2005. Common ancestry of herpesviruses and tailed DNA bacteriophages. *Journal of virology* **79**: 14967-14970.

574 Blewett EL, Sherrod CJ, Texier JR, Conrad TM, Dittmer DP. 2015. Complete Genome
575 Sequences of Mandrillus leucophaeus and Papio ursinus Cytomegaloviruses.
576 *Genome announcements* **3**.

577 Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A,
578 Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary
579 analysis. *PLoS computational biology* **10**: e1003537.

580 Brister JR, Ako-Adjei D, Bao Y, Blinkova O. 2015. NCBI viral genomes resource. *Nucleic*
581 *acids research* **43**: D571-577.

582 Brito AF, Pinney JW. 2017. Protein–Protein Interactions in Virus–Host Systems. *Frontiers*
583 *in microbiology* **8**: 1557.

584 Brito AF, Pinney JW. 2020. The evolution of protein domain repertoires: Shedding light on
585 the origins of the Herpesviridae family. *Virus Evol* **6**: veaa001.

586 Claramunt S, Cracraft J. 2015. A new time tree reveals Earth history's imprint on the
587 evolution of modern birds. *Science advances* **1**: e1501005.

588 Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. 2010. Jane: a new tool for the
589 cophylogeny reconstruction problem. *Algorithms for molecular biology : AMB* **5**: 16.

590 Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit
591 models of protein evolution. *Bioinformatics* **27**: 1164-1165.

592 Daugherty MD, Malik HS. 2012. Rules of engagement: molecular insights from host-virus
593 arms races. *Annual review of genetics* **46**: 677-700.

594 Davison AJ. 2002. Evolution of the herpesviruses. *Veterinary microbiology* **86**: 69-88.

595 Davison AJ. 2010. Herpesvirus systematics. *Veterinary microbiology* **143**: 52-69.

596 Davison AJ, Eberle R, Ehlers B, Hayward GS, McGeoch DJ, Minson AC, Pellett PE,
597 Roizman B, Studdert MJ, Thiry E. 2009. The order Herpesvirales. *Archives of*
598 *virology* **154**: 171-177.

599 de Vienne DM, Refregier G, Lopez-Villavicencio M, Tellier A, Hood ME, Giraud T. 2013.
600 Cospeciation vs host-shift speciation: methods for testing, evidence from natural
601 associations and relation to coevolution. *The New phytologist* **198**: 347-385.

602 dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PC, Yang Z. 2012. Phylogenomic
603 datasets provide both precision and accuracy in estimating the timescale of
604 placental mammal phylogeny. *Proceedings Biological sciences* **279**: 3491-3500.

605 Ehlers B, Dural G, Yasmum N, Lembo T, de Thoisy B, Ryser-Degiorgis MP, Ulrich RG,
606 McGeoch DJ. 2008. Novel mammalian herpesviruses and lineages within the
607 Gammaherpesvirinae: cospeciation and interspecies transfer. *Journal of virology* **82**:
608 3509-3516.

609 Escalera-Zamudio M, Rojas-Anaya E, Kolokotronis SO, Taboada B, Loza-Rubio E,
610 Mendez-Ojeda ML, Arias CF, Osterrieder N, Greenwood AD. 2016. Bats, Primates,
611 and the Evolutionary Origins and Diversification of Mammalian
612 Gammaherpesviruses. *mBio* **7**.

613 Feduccia A. 2003. 'Big bang' for tertiary birds? *Trends in Ecology & Evolution* **18**: 172-176.

614 Geoghegan JL, Duchene S, Holmes EC. 2017. Comparative analysis estimates the
615 relative frequencies of co-divergence and cross-species transmission within viral
616 families. *PLoS pathogens* **13**: e1006215.

617 George D, Blieck A. 2011. Rise of the earliest tetrapods: an early Devonian origin from
618 marine environment. *PloS one* **6**: e22136.

619 Gradstein FM, Ogg JG, Schmitz M, Ogg G. 2012. *The Geologic Time Scale 2012*. Elsevier,
620 The Amsterdam.

621 Hedges SB, Kumar K. 2009. *The Timetree of Life*. Oxford University Press, New York.

622 Hesse E, Buckling A. 2016. Host population bottlenecks drive parasite extinction during
623 antagonistic coevolution. *Evolution; international journal of organic evolution* **70**:
624 235-240.

625 Jackson A. 2005. The effect of paralogous lineages on the application of reconciliation
626 analysis by cophylogeny mapping. *Systematic biology* **54**: 127-145.

627 Johnson KP, Adams RJ, Page RD, Clayton DH. 2003. When do parasites fail to speciate
628 in response to host speciation? *Systematic biology* **52**: 37-47.

629 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
630 improvements in performance and usability. *Molecular biology and evolution* **30**:
631 772-780.

632 King AMQ, Lefkowitz EJ, Mushegian AR, Adams MJ, Dutilh BE, Gorbalenya AE, Harrach
633 B, Harrison RL, Junglen S, Knowles NJ et al. 2018. Changes to taxonomy and the
634 International Code of Virus Classification and Nomenclature ratified by the
635 International Committee on Taxonomy of Viruses (2018). *Archives of virology*
636 doi:10.1007/s00705-018-3847-1.

637 Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines,
638 Timetrees, and Divergence Times. *Molecular biology and evolution* **34**: 1812-1819.

639 Longdon B, Brockhurst MA, Russell CA, Welch JJ, Jiggins FM. 2014. The evolution and
640 genetics of virus host shifts. *PLoS pathogens* **10**: e1004395.

641 Lovisolo O, Hull R, Rosler O. 2003. Coevolution of viruses with hosts and vectors and
642 possible paleontology. *Advances in virus research* **62**: 325-379.

643 McGeoch DJ, Cook S, Dolan A, Jamieson FE, Telford EA. 1995. Molecular phylogeny and
644 evolutionary timescale for the family of mammalian herpesviruses. *Journal of*
645 *molecular biology* **247**: 443-458.

646 McGeoch DJ, Dolan A, Ralph AC. 2000. Toward a comprehensive phylogeny for
647 mammalian and avian herpesviruses. *Journal of virology* **74**: 10401-10406.

648 McGeoch DJ, Gatherer D. 2005. Integrating reptilian herpesviruses into the family
649 herpesviridae. *Journal of virology* **79**: 725-731.

650 McGeoch DJ, Rixon FJ, Davison AJ. 2006. Topics in herpesvirus genomics and evolution.
651 *Virus research* **117**: 90-104.

652 Page RDM, Charleston MA. 1998. Trees within trees: phylogeny and historical
653 associations. *Trends in Ecology & Evolution* **13**: 356-359.

654 Peters SE, McClennen M. 2015. The Paleobiology Database application programming
655 interface. *Paleobiology* **42**: 1-7.

656 Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular biology and*
657 *evolution* **25**: 1253-1256.

658 Raup DM. 1993. Extinction from a paleontological perspective. *European review* **1**: 207-
659 216.

660 Springer MS, Meredith RW, Gatesy J, Emerling CA, Park J, Rabosky DL, Stadler T,
661 Steiner C, Ryder OA, Janecka JE et al. 2012. Macroevolutionary dynamics and
662 historical biogeography of primate diversification inferred from a species
663 supermatrix. *PloS one* **7**: e49521.

664 Springer MS, Meredith RW, Janecka JE, Murphy WJ. 2011. The historical biogeography of
665 Mammalia. *Philosophical transactions of the Royal Society of London Series B,*
666 *Biological sciences* **366**: 2478-2502.

667 Thiry J, Keuser V, Muylkens B, Meurens F, Gogev S, Vanderplasschen A, Thiry E. 2006.
668 Ruminant alphaherpesviruses related to bovine herpesvirus 1. *Veterinary research*
669 **37**: 169-190.

670 Tischer BK, Osterrieder N. 2010. Herpesviruses--a zoonotic threat? *Veterinary*
671 *microbiology* **140**: 266-270.

672 Wertheim JO, Smith MD, Smith DM, Scheffler K, Kosakovsky Pond SL. 2014. Evolutionary
673 origins of human herpes simplex viruses 1 and 2. *Molecular biology and evolution*
674 **31**: 2356-2364.

676

677

FIGURES

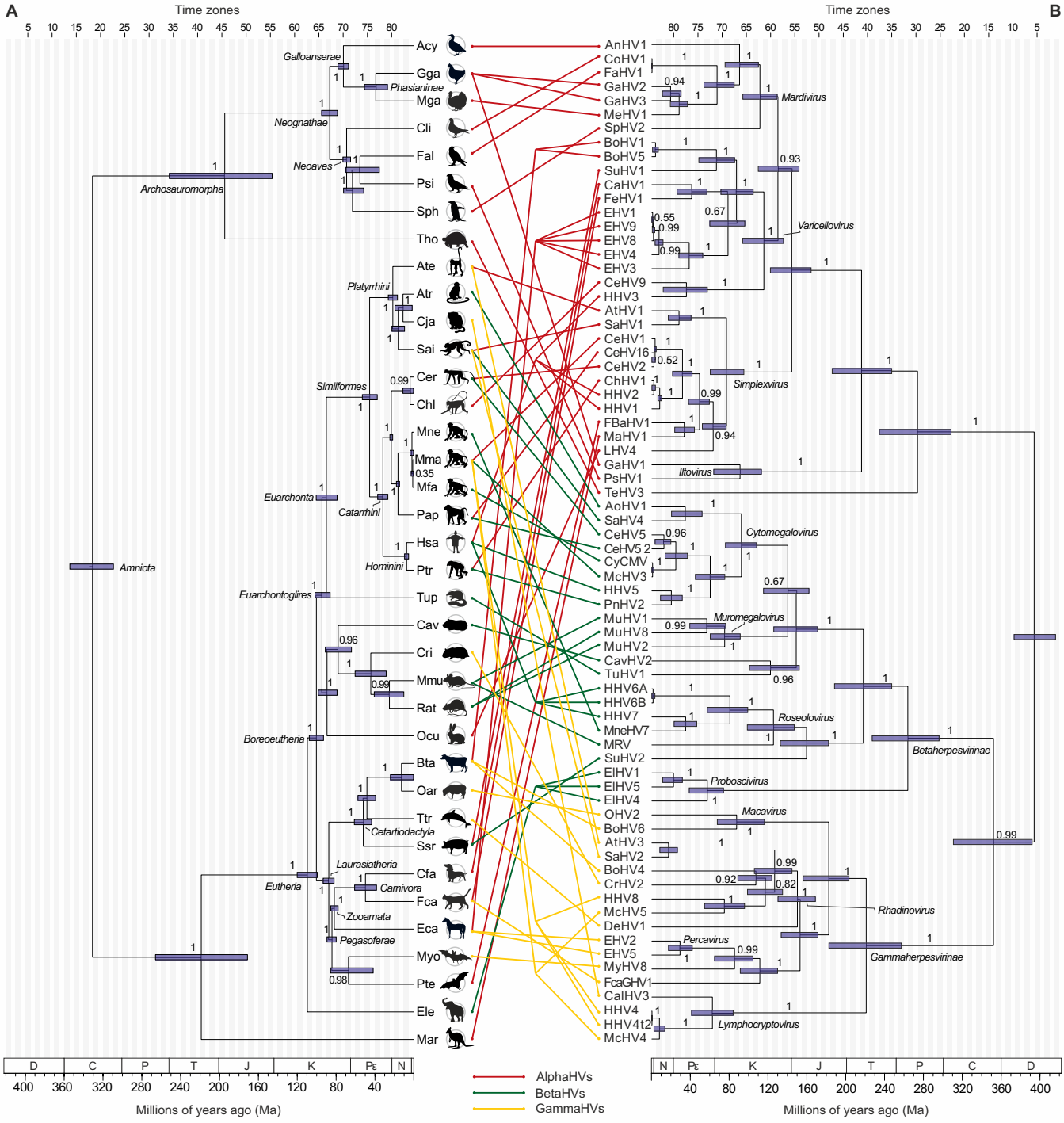


Figure 1

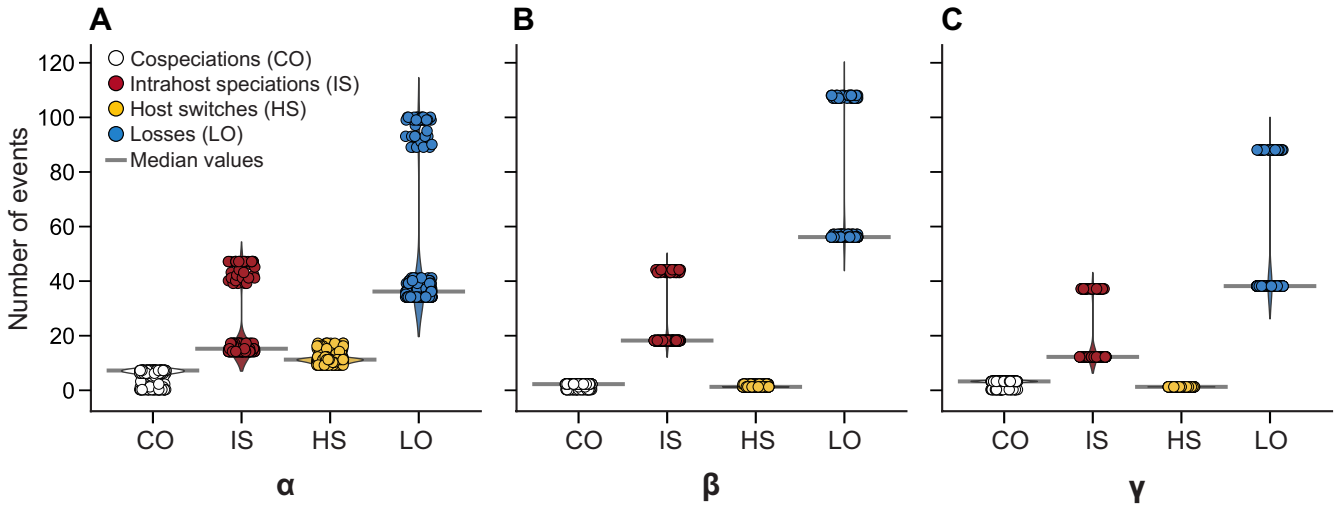
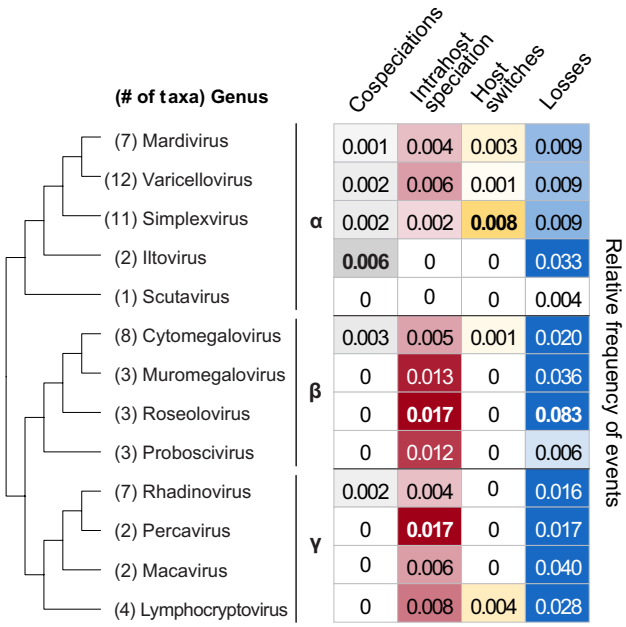

A
B
C
Cospeciations (CO)
Intrahost speciations (IS)
Host switches (HS)
Losses (LO)
Median values
Number of events
120
100
80
60
40
20
0
CO
IS
HS
LO
α
β
γ

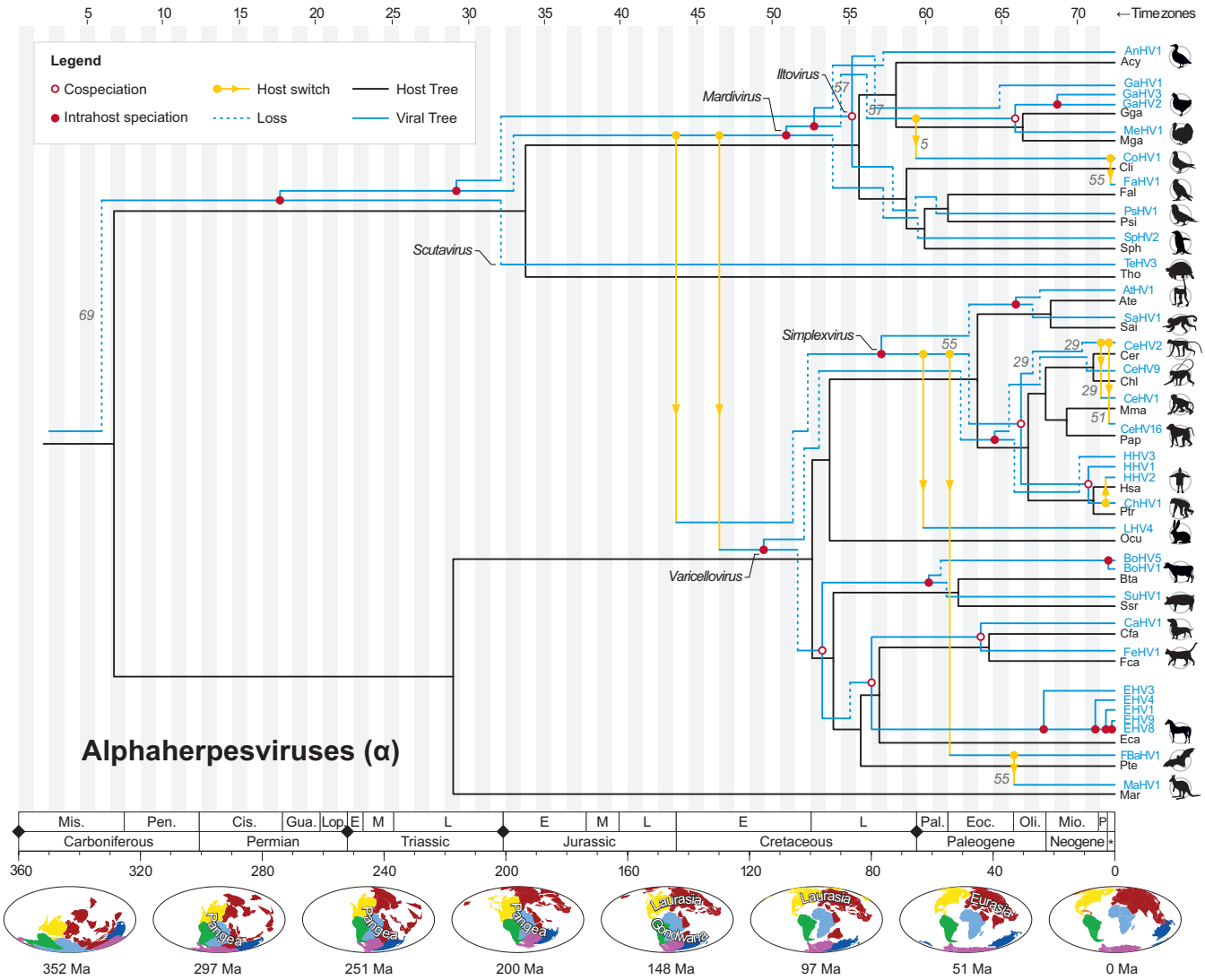
Figure 2



686

687 **Figure 3**

688



689

690

Figure 4

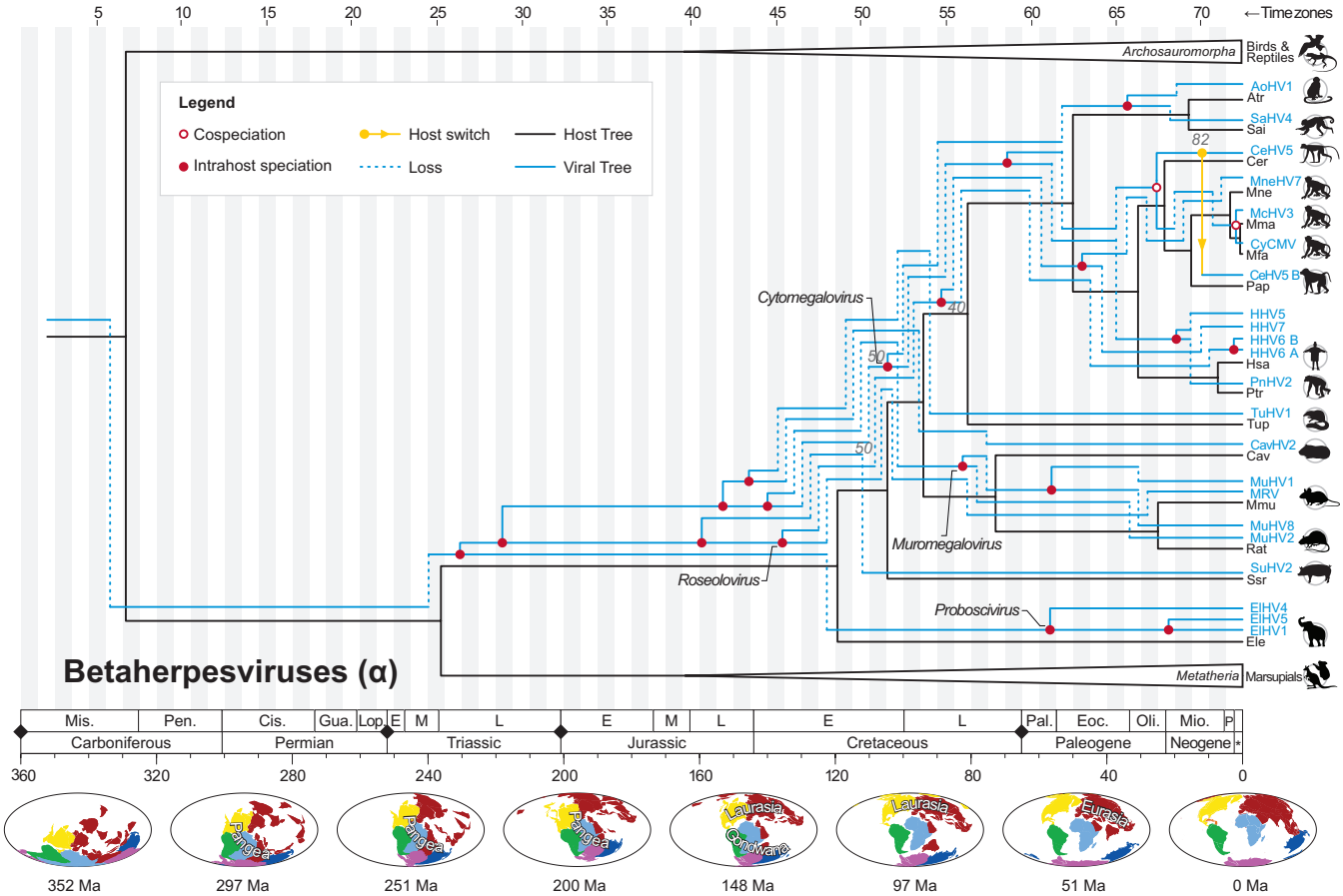

5
10
15
20
25
30
35
40
45
50
55
60
65
70
←Time zones
Legend
Cospeciation
Intrahost speciation
Host switch
Loss
Host Tree
Viral Tree
Archosauromorpha
Birds & Reptiles
AoHV1 Atr
SaHV4 Sai
CeHV5 Cer
MneHV7 Mne
McHV3 Mma
CyCMV Mfa
CeHV5 B Pap
HHV5
HHV7
HHV6 B
HHV6 A Hsa
PnHV2 Ptr
TuHV1 Tup
CavHV2 Cav
MuHV1
MRV Mmu
MuHV8
MuHV2 Rat
SuHV2 Ssr
ElHV4
ElHV5
ElHV1 Ele
Metatheria
Marsupials
Cytomegalovirus
Roseolovirus
Muromegalovirus
Proboscivirus
82
40
50
50
Betaherpesviruses (α)
Mis.
Pen.
Cis.
Gua.
Lop
E
M
L
E
M
L
E
L
Pal.
Eoc.
Oli.
Mio.
P
Carboniferous
Permian
Triassic
Jurassic
Cretaceous
Paleogene
Neogene
360
320
280
240
200
160
120
80
40
0
Pangea
Laurasia
Gondwana
Eurasia
352 Ma
297 Ma
251 Ma
200 Ma
148 Ma
97 Ma
51 Ma
0 Ma

Figure 5

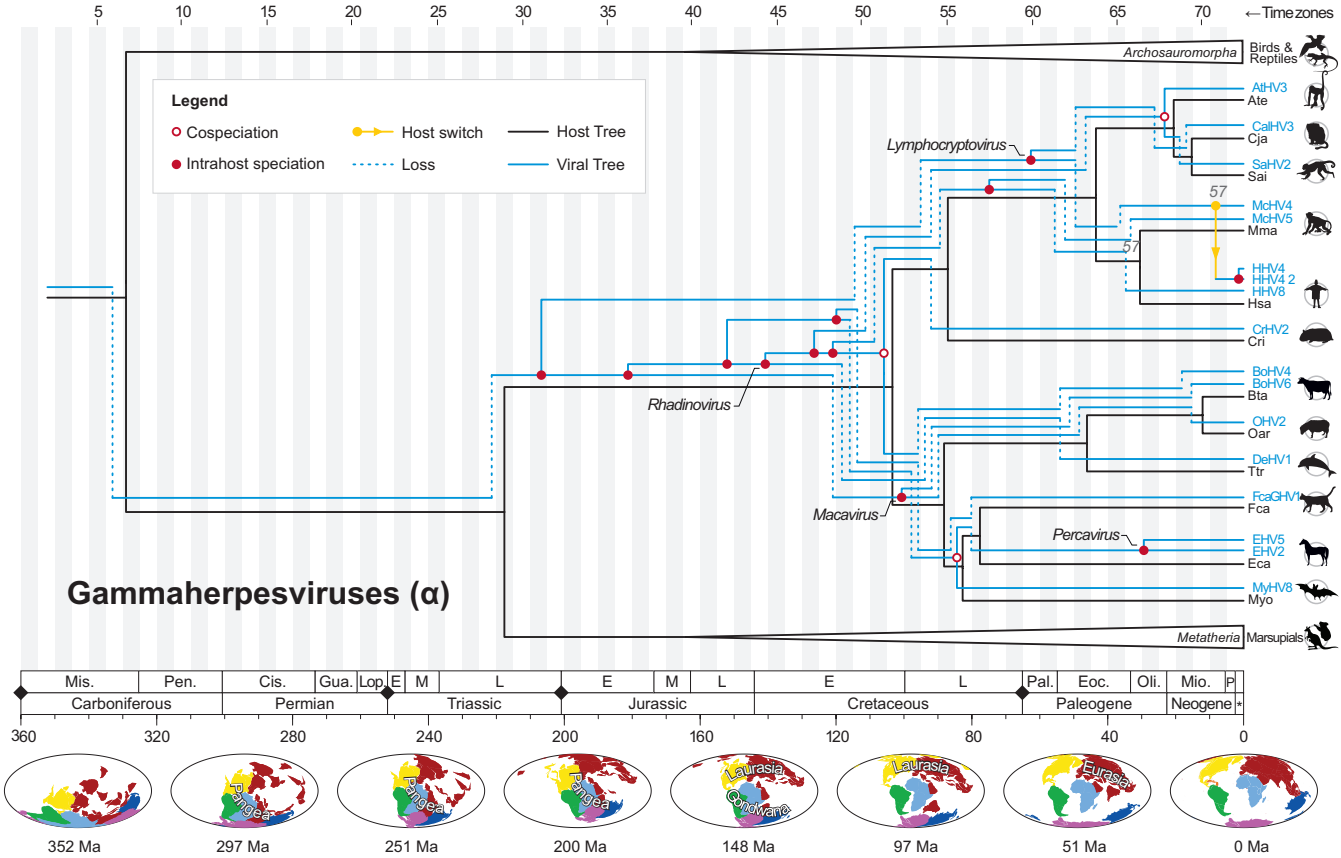

Legend
Cospeciation
Intrahost speciation
Host switch
Loss
Host Tree
Viral Tree
Gammaherpesviruses (α)
Time zones
Archosauromorpha
Birds & Reptiles
AtHV3
Ate
CalHV3
Cja
SaHV2
Sai
McHV4
McHV5
Mma
HHV4
HHV4 2
HHV8
Hsa
CrHV2
Cri
BoHV4
BoHV6
Bta
OHV2
Oar
DeHV1
Ttr
FcaGHV1
Fca
EHV5
EHV2
Eca
MyHV8
Myo
Metatheria
Marsupials
Lymphocryptovirus
Rhadinovirus
Macavirus
Percavirus
57
Mis.
Pen.
Cis.
Gua.
Lop.
E
M
L
Pal.
Eoc.
Oli.
Mio.
P
Carboniferous
Permian
Triassic
Jurassic
Cretaceous
Paleogene
Neogene
352 Ma
297 Ma
251 Ma
200 Ma
148 Ma
97 Ma
51 Ma
0 Ma

Figure 6