1 Ecological opportunity, radiation events and genomic innovations 2 shaped the episodic evolutionary history of papillomaviruses

3 Anouk Willemsen^{1,2*}, and Ignacio G. Bravo^{1,3}

4 ¹ Centre National de la Recherche Scientifique (CNRS), Laboratory MIVEGEC (CNRS IRD
5 Univ. Montpellier), Montpellier, France

6 ² Centre for Microbiology and Environmental Systems Science, University of Vienna,
7 Vienna, Austria

8 ³ Centre for Research on the Ecology and Evolution of Diseases (CREES), Montpellier,
9 France.

10 *Correspondence: anouk.willemsen@univie.ac.at

Abstract

Papillomaviruses (PVs) have a wide host range, infecting mammals, birds, turtles, and snakes. The recent discovery of PVs in different fish species allows for a more complete reconstruction of the evolutionary history of this viral family. In this study we perform phylogenetic dating to analyse evolutionary events that occurred during PV evolution, as well as to estimate speciation and evolutionary rates.

We have used four different data sets to explore and correct for potential biases that particular taxa combinations may introduce during molecular time inference. When considering the evolution of substitution rates we observed that short-term rate estimates are much higher than long-term rate estimates, also known as the time-dependent rate phenomenon. When considering the evolution of viral branching events (as a proxy for speciation rates), we show that these have not been constant through time, suggesting the occurrence of distinct evolutionary events such as adaptive radiations and/or changes in the available host niches. In a joint analysis with host speciation rates, we identified at least four different evolutionary periods, suggesting that the evolution of PVs has been multiphasic, and thus refining the previously suggested biphasic evolutionary scenario.

Thanks to the discovery of novel PVs in basal hosts and to the implementation of a time-dependent rate model for molecular dating, our results provide new insights into the evolutionary history of PVs. In this updated evolutionary scenario, ecological opportunity appears as one main driving force for the different radiation and key-innovation events we observe.

Keywords: virus evolution, molecular dating, time-dependent rate phenomenon, virus-host co-evolution and co-phylogeny, infection and cancer

34 Introduction

35 Papillomaviruses (PVs) are small non-enveloped circular dsDNA viruses with genome size
 36 between 6 and 8 kbp. The minimal PV genome consists of an upstream regulatory region
 37 (URR), an early gene region encoding for the E1 and E2 proteins, with in most cases the
 38 *E4* gene nested within *E2*, and a late gene region encoding for the L2 and L1 capsid
 39 proteins (García-Vallvé et al., 2005). Proteins in the early region are involved in viral
 40 replication and cell transformation, while the capsid proteins self assemble to yield virions
 41 and encapsidate the genome. During PV evolution, at different time points and in different
 42 lineages, the viral genomes have acquired, and in cases subsequently lost, the *E5*, *E6* and
 43 *E7* oncogenes (Van Doorslaer & McBride, 2016; Willemsen et al., 2019; Willemsen &
 44 Bravo, 2019). In chronic infections by certain viruses and in certain hosts, these genes are
 45 directly involved in the onset of cancer and behave thus as oncogenes.

46 PVs have been mainly described to infect mammals, but have also been found in reptiles
 47 and birds. More recently, PVs have also been discovered in fish, firstly isolated from a gilt-
 48 head sea bream (López-Bueno et al., 2016). Other PV genomes were later isolated from a
 49 rainbow trout, two haddocks, and a red snapper (Tisza et al., 2020). This discovery
 50 challenged our perspective on the origin of this viral family. Phylogenetic dating studies
 51 including this single fish PV genome date back the time to the most recent common
 52 ancestor (t_{mrca}) of PVs to 481 (95% HPD: 326–656; Van Doorslaer, Ruoppolo, et al.,
 53 2017) and 424 million years ago (95% HPD: 402–446; Willemsen & Bravo, 2019), and
 54 thus more ancient than previously thought.

55 The continuous discovery of novel PVs allows to add pieces to the puzzle of their
 56 evolutionary history. Notwithstanding, an enormous bias towards human taxon sampling
 57 remains, because certain human PVs (HPVs) are a major public health concern. Indeed,
 58 while the majority of PVs cause asymptomatic infections in skin and mucosa, a relatively

59 small number of oncogenic HPVs are associated to malignant lesions that can develop
60 into cancer (Monographs, 2012). For other animals, PVs associated to malignant lesions
61 have been found in a few polyphyletic lineages that mainly infect horses (Scase et al.,
62 2010), cows (Campo, 1997), rabbits (Kreider & Bartlett, 1981) and chamois (Mengual-
63 Chuliá et al., 2014).

64 Phylogenetic studies have revealed that virus-host codivergence is one of the main driving
65 forces of PV evolution: one third of the viruses' divergence patterns can be explained by
66 the hosts' divergence patterns (M. Gottschling et al., 2011). This match allows to identify
67 host divergence times that can be used to date the PV tree (Pimenoff et al., 2017; Rector
68 et al., 2007; Shah et al., 2010). Although virus-host codivergence plays an important
69 evolutionary role in PVs, it is not the only mechanism that has shaped PV diversification.
70 Other evolutionary processes such as recombination (Rector et al., 2008), lineage sorting
71 and host switches have also played a fundamental role in PV evolution, even in recent
72 times (Pimenoff et al., 2017). The best supported scenario proposes a biphasic evolution
73 (Félez-Sánchez et al., 2015; M. Gottschling et al., 2007, 2011), where an early PV
74 radiation would have generated the main extant viral crown groups, followed by
75 independent co-divergence between PVs and their hosts. Consequently, inconsistencies
76 between the virus and host trees are often detected, challenging the inference of ancestral
77 node ages of the PV tree.

78 Besides node ages, evolutionary rate is one of the key parameters used to characterise
79 the evolutionary history of viruses. Being dsDNA viruses, PVs belong within the group of
80 slow evolving viruses, with rates in the order of 10^{-7} to 10^{-9} nucleotide substitutions per site
81 per year (Pimenoff et al., 2017; Rector et al., 2007; Shah et al., 2010). These values are
82 several orders of magnitudes lower than the rates of their fast evolving counterparts, the

83 RNA viruses (Duffy et al., 2008). However, it is becoming widely accepted that the division
84 between viral types in terms of evolutionary rate is not as strict as historically assumed.
85 Over the last fifteen years it has become more evident that when performing molecular
86 dating, rate estimates may vary depending on the time frame of measurement, so that rate
87 estimates based on recent calibration nodes are much higher than those based on older
88 calibration nodes (Aiewsakun & Katzourakis, 2015, 2017; Duchene et al., 2014; Ho et al.,
89 2005, 2011). As a consequence, biased divergence times may be inferred, where long-
90 term rates tend to underestimate the divergence time, while short-term rates are prone to
91 overestimation. This time-dependent rate phenomenon has been detected in mitochondria
92 of different organisms (e.g. birds, fish, insects, penguins and primates Burrridge et al.,
93 2008; García-Moreno, 2004; Ho et al., 2005; Papadopoulou et al., 2010; Subramanian et
94 al., 2009), but has also been communicated for bacteria (Biek et al., 2015; Comas et al.,
95 2013; Feng et al., 2008; Rocha et al., 2006) as well as for multiple levels of viral taxonomy
96 (Aiewsakun & Katzourakis, 2015, 2016; Duchene et al., 2014; Gibbs et al., 2010). Despite
97 the differences in mean substitution rates between large viral groups (e.g. between DNA
98 and RNA viruses), the rate decay speed of time-dependent rates is reported to be
99 independent of viral taxonomy (Aiewsakun & Katzourakis, 2016; Duchene et al., 2014;
100 Gibbs et al., 2010). Therefore, corrections proposed to compensate for time-dependent
101 rates have proven to be useful in providing better estimation of the evolutionary time scale
102 of any virus (Aiewsakun & Katzourakis, 2015, 2017; Membrebe et al., 2019).

103 The time-dependency of molecular rate estimates is a near universal phenomenon and an
104 apparent artefact of the currently available reconstruction approaches. Nonetheless,
105 differences in rate measurements can also reflect biological processes. On the one hand,
106 high values for evolutionary rate, often recovered from short-term rate measurements, are
107 thought to approximate the spontaneous mutation rate (e.g. transient deleterious

108 mutations and transient short-sighted adaptations for the current host). On the other hand,
 109 low values for evolutionary rate, often recovered from long-term analyses, can better
 110 approximate the actual substitution rate (*i.e.* mutations that become fixed) over
 111 macroevolutionary timescales (Ho et al., 2011; Simmonds et al., 2019). However, this view
 112 does not explain the paradox of why viral genomes are conserved over the long-term while
 113 having an apparent unlimited evolutionary potential in terms of population size and
 114 mutation rate to evolve and adapt rapidly. Simmonds et al., 2019 proposed an alternative
 115 explanatory model where viral genome conservation is best explained by a niche-filling
 116 model in which fitness optimization is rapidly achieved in the viral hosts, and therefore viral
 117 long-term rates increasingly resemble those of their hosts.

118 In this study we revisit the evolutionary history of PVs with newly available ancestral fish
 119 PV genomes. We selected our calibration points and dated the PV tree using four different
 120 data sets based on the consistencies in virus-host codivergence patterns. Two of these
 121 data sets contain the virtually complete PV collection hitherto described, while the other
 122 two are different versions of reduced data sets to correct for the over-representation of
 123 humans and of other economically important hosts. First, we investigate whether a time-
 124 dependent rate model better fits PV evolution data than previously used models. Second,
 125 we compare evolutionary rates, diversification rates, and node age estimates among the
 126 different PV data sets. In addition, we compare the PV evolutionary- and diversification
 127 rate estimates with those of the corresponding hosts. Based on our observations we
 128 propose an updated evolutionary scenario and point out important events that occurred
 129 during the evolution of PVs.

130 **Materials and Methods**

131 *Data collection and alignments*

For this study, 359 full length PV genomes were downloaded from the PaVE (<https://pave.niaid.nih.gov/>, Van Doorslaer, Li, et al., 2017) and GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) databases (Table S1). The *E1*, *E2*, *L2* and *L1* genes were extracted and aligned individually at the amino acid level using MAFFT v.7.271 (Katoh & Standley, 2013), corrected manually, and backtranslated to nucleotides using PAL2NAL v.14 (Suyama et al., 2006). The alignment was filtered using Gblocks v.0.91b (Castresana, 2000), with the following parameters: -t=c, -b1=50% of the number of sequences + 1, -b2=50% of the number of sequences + 1, -b3=8, -b4=3, -b5=a, -b0=3. For tree construction, *E1*, *E2*, *L2* and *L1* were concatenated using a custom perl script.

141 *Prior phylogenetic analyses*

To detect recombinant taxa as well as incongruent taxa between the early gene tree and the late gene tree, we used the filtered concatenated early genes (*E1-E2*) and late genes (*L2-L1*). For these two alignments, the LG+I+ Γ protein substitution model was identified as the best suited model using ProtTest v.3.4.2 (Darriba et al., 2011). ML-based phylogenetic analyses were conducted using RAxML v.8.2.9 (Stamatakis, 2014) under the GTR+ Γ 4 model for the nucleotide alignment using six partitions (three for each gene corresponding to each codon position), under the LG+I+ Γ model for the amino acid alignment using two partitions (one for each gene), and using 1000 bootstrap replicates. The trees were rooted using the SaPV1 (*Sparus aurata papillomavirus 1*) sequence (López-Bueno et al., 2016; Fig. S1-S2). Rogue taxa were identified using the algorithm implemented in RAxML (Pattengale et al., 2011). The topologies of the *E1-E2* and *L2-L1* trees were compared using the disagree method in TOPD-fMtS v.3.3 (Puigbo et al., 2007), which allowed us to identify the disagreeing taxa between the two tree topologies.

155 The previously identified recombinant PVs isolated from Cetaceans (PphPV1-2, TtPV1-7,
156 DdPV1, PsPV1) (Marc Gottschling et al., 2011; Rector et al., 2008; Robles-Sikisaka et al.,
157 2012) also showed to confidently disagree in position between the early and late gene
158 trees in our analyses. Therefore, these recombinant PVs were removed for generating the
159 Full Data set (FD), leaving us with a data set of 343 PV genomes (of which 200 infect
160 humans). Based on the taxa that agree between the *E1-E2* and *L2-L1* trees, we generated
161 two different versions of reduced data sets containing representatives for each PV species
162 and PV type: Representative Data set 1 (RD1) and Representative Data set 2 (RD2). RD1
163 and RD2 both contain 130 PVs (of which 48 infect humans), and share 85 terminal taxa.
164 RD1 and RD2 were constructed to compensate for certain over-represented host taxa in
165 the FD. For example, the 200 human PVs and the 21 bovine PVs present in the FD were
166 respectively represented by 48 and 6 terminal taxa in the reduced data sets (see [Table](#)
167 [S1](#)). For the sake of clarity we would like to stress that classification within
168 *Papillomaviridae* is explicitly based on genetic distance, where each PV type is a unique
169 genomic entity. During the establishment of the boundaries for defining the different
170 taxonomic categories within the *Papillomaviridae* (genera – species – type – variant) it was
171 recognized that distribution of pairwise genetic distances is multimodal (De Villiers et al.,
172 2004). The PV working group within the International Committee on the Taxonomy of
173 Viruses decided then to follow these “natural categories” to delineate the boundaries of the
174 taxonomic levels, leading to an official definition of phylogenetic taxa at higher taxonomic
175 levels, as an ensemble of PV types grouped into genera, species and subfamilies (Bernard
176 et al., 2010; Burk et al., 2013; Bzhalava et al., 2015; De Villiers et al., 2004; Van Doorslaer
177 et al., 2018; see also <https://talk.ictvonline.org/taxonomy/>).

178 Our FD, RD1 and RD2 contained one single fish PV genome (SaPV1), but during the
179 study four other fish PV genomes were made available in GenBank (accessions:

MH510267, MH616908, MH617143, MH617579; Tisza et al., 2020). As the sequences of these new PVs add relevant information to the basal clade of the tree, these genomes were added to the FD, and this new data set containing 347 PV genomes was named FDF (FD + additional fish PVs).

For all four data sets (FDF, FD, RD1, and RD2), the concatenated *E1-E2-L2-L1* alignments were used to construct ML trees with RAxML v.8.2.9 under the GTR+ Γ 4 model for the nucleotide alignments, using twelve partitions (three for each gene corresponding to each codon position), under the LG+I+ Γ model for the amino acid alignments using four partitions (one for each gene), and in both cases using 1000 bootstrap replicates (Fig. S3-S6).

Phylogenetic time inference

Based on the ML constructed trees, 18 calibration nodes were selected on subtrees where the *E1-E2* and *L2-L1* trees did not show discrepancies (Fig. S1-S2) and where the host tree matched the PV tree. The host time tree (Fig. S7) was recovered from TimeTree (<http://www.timetree.org/>; Kumar et al., 2017), based on the list of all known PV host species included in this study. Calibration times were based on host divergence molecular clock estimates collected from TimeTree, integrating the corresponding confidence intervals in the prior (Table 1). The effect of the calibration nodes, and therewith forced clades, on the topology of the tree was validated by constructing ML trees constrained to the calibrations used and subsequent comparison to the corresponding unconstrained tree using a Shimodaira-Hasegawa test (Shimodaira & Hasegawa, 1999) as implemented in RAxML (Stamatakis, 2014). At the nucleotide level, the constrained trees did not test significantly worse than the unconstrained trees (Table S2). At the amino acid level, the constrained FDF and RD2 trees tested significantly worse. The nucleotide-based

204 constrained ML trees were used as a starting trees for time inference at the nucleotide
205 level.

Number	Calibration node	Divergence estimate (million years ago)	CI (million years)
1	Root	435	425 - 446
2	Aves/Testudines	254	240 - 268
3	Perissodactyla/Cetartiodactyla	78	74 - 81
4	Bovidae1/Cervidae1	27.31	23.13 - 31.49
5	Bovidae2/Cervidae2	27.31	23.13 - 31.49
6	Laurasiatheria/Primates	96	91 - 102
7	Platyrrhini/Catarrhini	43.20	40.60 - 45.70
8	Homo/Colobus	29.44	27.61 - 31.28
9	Homo/Pan	6.65	6.23 - 7.07
10	Pan troglodytes/Pan paniscus	2.82	2.40 - 3.24
11	Rodentia/Primates	90	85 - 94
12	Laurasiatheria/Euarchontoglires	96	91 - 102
13	Insectivora/Carnivora	89	83 - 96
14	Hyaenidae/Felidae	40	33 - 46
15	Primates/Glires	90	85 - 94
16	Bovidae3/Cervidae3	27.31	23.13 - 31.49
17	Feliformia/Caniformia	54	52 - 57
18	Cricetidae/Muridae	32.70	27.90 - 37.40

Table 1. Calibration points used for molecular dating of the PV tree. The estimates and the corresponding confidence intervals (CIs) are based on host divergence estimates (TimeTree: <http://www.timetree.org/>). Numbers in the first column correspond to those used in subsequent graphs and tables in this article to refer to the calibration nodes.

206 Before a time-dependent rate model was implemented in BEAST (Membrebe et al., 2019;
207 Suchard et al., 2018), we manually verified whether a time-dependency of molecular rate
208 estimates exists for PVs. Bayesian time inference was performed at the nucleotide level
209 using BEAST v.1.8.3 (Alexei J Drummond et al., 2012), under the GTR+ Γ 4 model, using
210 twelve partitions, the uncorrelated relaxed clock model (Alexei J Drummond et al.,
211 2006) with a lognormal distribution and a continuous quantile parametrisation (Li &

Drummond, 2012), and the Yule speciation process tree prior (Gernhard, 2008; Yule, 1925). For each calibration node, a prior was defined with a normal distribution around the host divergence estimate and with a standard deviation based on the confidence intervals indicated in [Table 1](#). Time inference was performed separately for each calibration node, on both RD1 and RD2. Thus, per data set, 18 independent time inferences using single calibration nodes were performed. A further description of the manual verification and following correction of time-dependency of molecular rate estimates can be found in [Supplementary File 1](#).

Subsequently, Bayesian time inference was performed at the nucleotide level using BEAST v.1.10.4 (Suchard et al., 2018), under the GTR+ Γ 4 model, using four partitions (one for each gene). Using data sets RD1 and RD2, four different clock models were tested under the Yule speciation process tree prior (Gernhard, 2008; Yule, 1925): (i) the strict clock model, (ii) the uncorrelated relaxed clock model (Alexei J Drummond et al., 2006) with a lognormal distribution and a continuous quantile parametrisation (UCLD; Li & Drummond, 2012), (iii) a uniform time-dependent rate (TDRuni) clock model, and (iv) an exponential time-dependent rate (TDRexp) clock model (Membrebe et al., 2019). The strict clock model assumes a single substitution rate for the whole tree. The UCLD assumes that that substitution rate along each branch is drawn independently from a lognormal distribution, nonetheless one single mean substitution rate is calculated over the tree. The TDR clock model allows for substitution rate variation over time. The TDRuni and TDRexp epoch models are set up as described in Membrebe et al., 2019, with custom uniform time intervals (with boundaries $0 < 10 < 20 < \dots < 400 < \infty$) and custom exponentially distributed time intervals (with boundaries $0 < 10^{-5} < 10^{-4} < \dots < 10^2 < \infty$), that are expected to cover the depth of the PV phylogeny. The Yule speciation process branching model assumes a constant speciation rate with no extinction. The clock model with the best model fit was

also tested with a coalescent Bayesian Skyline tree prior (A. J. Drummond et al., 2005), that allows the population size to vary stochastically over time. For each calibration node, a normal distribution was assumed with the host divergence estimate and standard deviation based on the confidence intervals indicated in (Table 1). The standard deviation of the root was relaxed to 50 million years. For each reduced data set, and each model, four independent MCMC chains were run for a maximum of 10^7 generations, sampling every 10^4 . We compared the model fits of the different clock models with the (log) marginal likelihood estimates obtained using stepping-stone sampling (Baele et al., 2012, 2013), with 100 paths steps, and a chain length of 10^6 . From all models tested with the RD1 and RD2 data sets, the best model was also used for time inference with the FD and FDF data sets. For each of these, seven independent MCMC chains were run for a maximum of 3×10^7 generations, sampling every 3×10^4 .

Statistics and graphics

Statistical analyses and graphics were done using R (R Core Team, 2014), with the aid of the packages "ape", "car", "dplyr", "ggfortify", "ggplot2", "ggtree", "lawstat", "overlapping", "pgirmess", "reshape", "stats", and "strap". Computation of optimal breakpoints in the LTT plots was performed using the R package "strucchange". The optimal number of breakpoints was first calculated using the default 'strucchange' parameters. From this analyses the optimal number of breakpoints for the host LTT plot was found to be 3. Since the older nodes contain very few observations (due to limited sampling efforts in basal PVs and thus a limited number of basal branching events), we refined the analyses by specifying a minimal sample size of 3 data points for each segment (parameter h), together with a maximal number of breaks to be calculated set at either 3, 4 or 5. From these analyses, a model with 3 breakpoints (at 148, 64 and 10 Ma) resulted to best fit the

261 phylogenetic reconstruction data of the hosts. The final display of the graphics was
 262 designed using Inkscape v.0.92 (<https://inkscape.org/en/>). The silhouettes in Fig. 1 were
 263 obtained from Freepik (<https://www.freepik.com/>).

264 **Results**

265 *PV crown-groups form well-supported clades, albeit with unclear relative positions among*
 266 *them*

267 In this study we use four different data sets, with a less exhaustive (RD1 and RD2) and a
 268 more exhaustive (FD and FDF) representation of all PV genomes in the databases (see
 269 Materials and Methods). As some PV lineages —mainly PVs infecting humans and some
 270 economically important species such as cattle, dogs and horses— are over-represented in
 271 the full data sets (FD and FDF), we chose to also work with two different reduced
 272 representative data sets (RD1 and RD2). For the construction of RD1 and RD2, the full
 273 data sets were used to select one representative for each PV species and type. We would
 274 like to stress that for PVs, a type describes a unique genomic entity, genetically different
 275 from and within defined boundaries of nucleotide identity to other sister taxa (see also
 276 Materials and Methods). RD1 contains representative species with basal and often shorter
 277 branches as compared to the derived taxa with often longer branches chosen for RD2.
 278 Thus for RD2, more diverse PV genomes are being compared.

279 For all data sets, the early and the late genes were concatenated (*E1-E2-L2-L1*) and trees
 280 were constructed at the nucleotide and amino acid level under a maximum likelihood
 281 framework. Overall, we observed well-supported clades for the different PV crown groups
 282 (Fig. 1 and Fig. S3-S6): Alpha-OmikronPVs, Beta-XiPVs, Lambda-MuPVs, and Delta-
 283 ZetaPVs. Despite the conserved support of these large clades, their relative positions,

284 however, vary for the different trees constructed at the nucleotide and amino acid levels,
285 as well as between the different data sets.

286 *A time-dependent rate model describes best the evolutionary history of PVs*

287 When performing time inference using one single calibration at a time ([Table 1](#)), we
288 observed that a time-dependency of molecular rate estimates exists for PVs, with younger
289 calibrations rendering higher inferred values of the molecular evolutionary rate. As shown
290 in [Fig. 2](#), there is a strong correlation between the substitution rate and the inferred time
291 ([Fig. 2a](#): RD1, *Spearman's rho* = 0.8225, *S* = 1766, *p* = 2.419e-05; [Fig. 2b](#): RD2,
292 *Spearman's rho* = 0.5170, *S* = 1470, *p* = 0.0299). For RD1 the substitution rate inferred for
293 the youngest node (2.62 Ma: 1.66×10^{-8} s/s/y) is 3.3 times higher than the substitution rate
294 inferred for the root of the tree (434.11 Ma: 5.03×10^{-9} s/s/y, see also [Table S3](#)). For RD2
295 the substitution rate inferred for the youngest node (3.61 Ma: 0.68×10^{-8} s/s/y) is 1.3 times
296 higher than the substitution rate inferred for the root of the tree (434.19 Ma: 5.22×10^{-9} s/s/y,
297 see also [Table S3](#)). The difference in substitution rate for the younger nodes is probably
298 related to the criteria for taxa choice in both reduced data sets, which explore different
299 terminal branch lengths and thus evolutionary distances for recent nodes, with a more
300 limited impact on the values inferred for deeper nodes. When performing time inference
301 using all 18 calibrations nodes together, the variation of evolutionary rates over time is not
302 captured under the uncorrelated relaxed clock model (UCLD). Therefore, we have
303 estimated our node ages by correcting for the time-dependent rate phenomenon by fitting
304 a power-law rate decay model to the single calibration data (as described in
305 [Supplementary File 1](#)). Recently, a time-dependent rate (TDR) clock model was
306 implemented in BEAST, that accommodates rate variation through time (Membrebe et al.,
307 2019). When performing time inference using all 18 calibration nodes ([Table 1](#)) and the

308 TDR clock model, we observe a strong TDR effect under an epoch structure with both
 309 uniformly and exponentially distributed time intervals (Fig. S8). To investigate whether the
 310 TDR model indeed better describes the evolutionary history of PVs, we tested the TDR
 311 clock model against the UCLD and strict clock model. Both the exponential and uniform
 312 TDR clock models yield considerably higher log marginal likelihood estimates compared to
 313 the UCLD and strict clock models (Table 2). The exponential TDR model shows the best fit
 314 to the data among the clock models being compared. Fig. 3a depicts the molecular rate
 315 estimates under the exponential TDR model with a pronounced rate increase towards the
 316 present for all four data sets. Comparing all four data sets under the exponential TDR
 317 model, we estimate the short-term PV rate to be between 1.29×10^{-8} and 1.42×10^{-8}
 318 substitutions/site/year (s/s/y), while the long-term PV rate is estimated to be between
 319 2.42×10^{-9} and 4.58×10^{-9} s/s/y (Table 3). The evolutionary rate estimates for all epochs can
 320 be found in Table S4. Under the uniform TDR model and using the reduced data sets, the
 321 estimated short-term PV rate (Table S5, between 0-10 Ma: $\sim 1.39 \times 10^{-8}$ s/s/y) falls within the
 322 range of the rate obtained under the exponential TDR model. The long-term PV rate is
 323 estimated to be lower under the uniform TDR model (Table S5, above 400 Ma: $\sim 1.81 \times 10^{-9}$
 324 s/s/y), most probably related to the last epoch starting 300 Ma later compared to the last
 325 epoch under the exponential TDR model (above 100 Ma). Not unexpected, the estimated
 326 short-term PV rates are remarkably higher while the estimated long-term PV rates are
 327 lower, than the overall evolutionary rate estimate of $\sim 8 \times 10^{-9}$ s/s/y inferred under the UCLD
 328 model (Table 3). With the reduced data sets we infer higher short-term and lower long-term
 329 rate as compared to the full data sets (Fig. 3a), i.e. a stronger time-dependent rate effect,
 330 as is evident in the regression coefficients for RD1 and RD2 and the regression
 331 coefficients for FD and FDF (Table 2).

Data set	Clock model	Tree prior	Log P(data M)	Parameters	Mean	95% HPD
RD1	TDRexp	Yule Process	-420158.160	Intercept β_0	-3.600	(-3.692; -3.500)
				Slope β_1	-0.384	(-0.418; -0.355)
	TDRexp	Bayesian Skyline	-438793.035	Intercept β_0	-3.657	(-3.761; -3.556)
				Slope β_1	-0.365	(-0.396; -0.333)
	TDRuni	Yule Process	-452190.114	Intercept β_0	-3.554	(-3.670; -3.439)
				Slope β_1	-0.423	(-0.463; -0.385)
RD2	UCLD	Yule Process	-5.842e+300	Mean μ	0.949	(0.918; 0.980)
				Dispersion σ	0.235	(0.207; 0.264)
	Strict	Yule Process	-5.774e+300	Clock rate r	1	NA
	TDRexp	Yule Process	-428278.860	Intercept β_0	-3.778	(-3.877; -3.669)
				Slope β_1	-0.326	(-0.359; -0.294)
RD2	TDRexp	Bayesian Skyline	-449076.075	Intercept β_0	-3.837	(-3.952; -3.731)
				Slope β_1	-3.087	(-0.338; -0.271)
	TDRuni	Yule Process	-445062.696	Intercept β_0	-3.737	(-3.868; -3.607)
				Slope β_1	-0.364	(-0.406; -0.321)
	UCLD	Yule Process	-6.339e+300	Mean μ	0.951	(0.920; 0.984)
				Dispersion σ	0.238	(0.209; 0.269)
FD	TDRexp	Yule Process	NA	Clock rate r	1	NA
FDF	TDRexp	Yule Process	NA	Intercept β_0	-3.970	(-4.063; -3.887)
				Slope β_1	-0.224	(0.250; -0.194)
FDF	TDRexp	Yule Process	NA	Intercept β_0	-3.928	(-4.019; -3.840)
				Slope β_1	-0.238	(-0.265; -0.208)

Table 2. Model fit and regression coefficients for the time-dependent rate clock models. For RD1 and RD2 the model fit of different clock and tree models using (log) marginal likelihood estimates obtained using stepping-stone sampling is given. The exponential time-dependent clock model (TDRexp), the uniform time-dependent clock model (TDRuni), the uncorrelated relaxed clock model (UCLD) and the strict clock mode were compared. The TDRexp model with a Yule tree prior yields the highest log marginal likelihood estimate for both data sets. For RD1, RD2, FD and FDF, the regression coefficients of the TDRexp model were compared. The steeper slopes for RD1 and RD2 indicate a stronger time-dependent rate effect compared to FD and FDF (see also Fig. 3a).

Data set	Nr. of taxa	Mean rate (s/s/y)	95% HPD	Short-term rate (s/s/y) (1 - 10 Ma)	95% HPD	Long-term rate (s/s/y) (above 100 Ma)	95% HPD
RD1	130	7.90×10^{-9}	(7.45×10^{-9} ; 8.35×10^{-9})	1.42×10^{-8}	(1.27×10^{-8} ; 1.56×10^{-8})	2.42×10^{-9}	(2.16×10^{-9} ; 2.68×10^{-9})
RD2	130	8.01×10^{-9}	(7.55×10^{-9} ; 8.47×10^{-9})	1.31×10^{-8}	(1.18×10^{-8} ; 1.45×10^{-8})	2.92×10^{-9}	(2.60×10^{-9} ; 3.24×10^{-9})
FD	343	7.99×10^{-9}	(7.62×10^{-9} ; 8.36×10^{-9})	1.29×10^{-8}	(1.16×10^{-8} ; 1.41×10^{-8})	4.58×10^{-9}	(4.18×10^{-9} ; 5.06×10^{-9})
FDF	347	7.68×10^{-9}	(7.31×10^{-9} ; 8.02×10^{-9})	1.31×10^{-8}	(1.19×10^{-8} ; 1.47×10^{-8})	4.39×10^{-9}	(4.00×10^{-9} ; 4.86×10^{-9})

Table 3. The mean, short- and long-term evolutionary rate estimates for PVs. The mean evolutionary rates were inferred by using the uncorrelated relaxed clock model (UCLD) with a Yule tree prior and using all 18 calibration nodes together. The short- and long-term evolutionary rates were inferred by time inference using the exponential TDR model with a Yule tree prior and using all 18 calibration nodes together. Values are in substitutions per site per year (s/s/y).

332 *PVs have not diversified at a constant rate*

333 The dated phylogenetic trees using RD1, RD2, FD, and FDF, with an exponential TDR
334 model and a Yule tree prior are shown in [Fig. 3b](#) and [Fig. S9-12](#). A detailed annotated
335 version of the tree obtained using RD2 is depicted in [Fig. 1](#). The main differences between
336 the reduced and full data sets are observed in the ancestral nodes, where those of RD1
337 and RD2 are inferred to be much older than those of FD and FDF ([Fig. 3b](#)). To visualize
338 the timing of PV diversification, we constructed lineages-through-time (LTT) plots based on
339 the dated trees and compared these to the host tree ([Fig. 3c](#)). For both viral and host LTT
340 plots we observe different evolutionary periods, where the branching events are not
341 constant through time, suggesting the occurrence of distinct evolutionary events such as
342 adaptive radiations. For viral data we can distinguish four evolutionary periods (divided by
343 grey vertical bars in [Fig. 3c](#)), that are based on the observation of breakpoints in the four
344 viral LTT plots (RD1, RD2, FD and FDF). Using the host LTT plot, we computed the
345 optimal breakpoints, allowing us to establish general borders (at 148, 64 and 10 Ma) for
346 four evolutionary periods (divided by black dashed vertical lines in [Fig. 3c](#)). Our results are

347 consistent with the main periods proposed for the evolution of the lineage leading to
 348 placental mammals: (i) 148 Ma (95% CI: 110.0-158.6) could span the divergence time of
 349 Placentals and Marsupials and the initial basal diversification within mammalian orders; (ii)
 350 64 Ma (95% CI: 62.0-66.8) corresponds to intra-ordinal the diversification of placental
 351 mammals; and (iii) 10 Ma (95% CI: 7.7-9.7) corresponds to a probable much slower rate of
 352 diversification until present. These boundaries inferred based on the host tree roughly
 353 match the borders for the observed viral evolutionary periods, and were used to compare
 354 the slopes of the viral and host LTT plots (Fig. 3d and Table 4). Our analysis of the LTT
 355 slopes indicates that the rate of emergence of novel viral lineages was higher than the rate
 356 of novel host lineages in periods II and III (*i.e.* steeper slopes for the emergence of PV
 357 lineages). It is important to note that our estimates increase in uncertainty as we approach
 358 the root of our tree (period I), essentially due to the limited number of viral taxa retrieved
 359 from non-mammalian hosts.

360 In period I (between ~600 and ~148 Ma) we first observe a slow increase of PV lineages
 361 that corresponds to the basal PV clades: PVs infecting fish, birds, turtles and ancestral
 362 mammals. In period II (between ~148 and ~64 Ma), the number of PV lineages surpasses
 363 the number of host lineages, growing at a ~1.5 times faster rate. We interpret that
 364 evolutionary period II corresponds to a viral radiation, that goes in parallel with the crown
 365 radiation of placental mammals ~100 Ma, that generated a large number of viral lineages.
 366 In period III (between ~64 and ~10 Ma) both the PV and host slopes decrease,
 367 notwithstanding that the number of PV lineages continues to increase at a rate that is ~1.8
 368 times faster as compared to the hosts. We interpret that period III corresponds to the
 369 diversification of placental mammals within the host crown-groups and the parallel
 370 diversification of PVs with their hosts. Lastly, in period IV (between ~10 Ma and the
 371 present) we observe a further decrease in both PV and host slopes. We interpret that in

addition to virus host co-evolution, the flattening slopes in periods III and especially IV, indicate a much slower diversification of PVs within their hosts (Fig. 3d and Table 4). The lower slope values for PV diversification in the reduced data sets compared to the full data sets arises most likely from the reduction of the number of terminal taxa associated to individual hosts.

The LTT plots and the LTT slope analysis suggest that the per-lineage speciation rates have not remained constant through time. Thus we can reject the null hypothesis of constant diversification rates on the FDF, FD, RD1, and RD2 data sets, suggesting that the observed trends in PV number of lineages through time is compatible with evolutionary events such as adaptive radiations and/or key adaptations. The Yule tree prior used for time inference assumes a constant speciation rate with no extinction, indicating that this tree prior is an inadequate model. Therefore, we also performed time inference with a Bayesian Skyline tree prior (BS) and the exponential TDR clock model. When comparing the LTT plots of trees inferred with either a Yule or a BS tree prior, we observed that these render highly similar results (Fig. S13). Nevertheless, time inference with a Yule tree prior yields a higher log marginal likelihood estimate than the inference with the BS tree prior (Table 2). This probably indicates that the concatenated PV gene alignments and the calibration nodes are much more informative than the tree prior, which plays thus a minor role on the results.

Period	Data set	Slope	SE	df	F-stat.	adj. R ²	p-value
I 148-600 Ma	RD1	1.1239	0.1736	2	41.93	0.932	0.0230
	RD2	1.1560	0.1988	2	33.8	0.916	0.0283
	FD	0.9801	0.2562	1	14.64	0.872	0.1627
	FDF	1.3909	0.3527	2	15.56	0.829	0.0587
	hosts	1.7320	0.2027	5	72.98	0.923	0.0004
II 64-148 Ma	RD1	3.0603	0.1057	26	838.5	0.969	< 2.2e-16
	RD2	3.2792	0.0815	29	1618	0.982	< 2.2e-16

		FD	3.8183	0.1025	37	1388	0.973	< 2.2e-16
		FDF	3.6887	0.0937	41	1550	0.974	< 2.2e-16
		hosts	2.2782	0.0855	16	710	0.977	1.104e-14
III	10-64 Ma	RD1	0.8422	0.0286	85	865	0.910	< 2.2e-16
		RD2	0.7392	0.0361	85	419.4	0.830	< 2.2e-16
		FD	0.9082	0.0137	223	4385	0.951	< 2.2e-16
		FDF	0.8594	0.0128	219	4492	0.953	< 2.2e-16
		hosts	0.4562	0.0091	38	2526	0.985	< 2.2e-16
IV	0-10 Ma	RD1	0.0315	0.0083	8	14.46	0.599	0.0052
		RD2	0.0298	0.0070	5	18.21	0.742	0.0080
		FD	0.0759	0.0112	73	45.88	0.378	2.736e-09
		FDF	0.0798	0.0111	76	51.76	0.397	3.769e-10
		hosts	0.1442	0.0097	13	221.4	0.940	1.527e-09

Table 4. The slopes of the LTT lines in Fig. 3c. For each data set (RD1, RD2, FD and FDF) the slope is given for periods I-IV and also shown in Fig. 3d. These periods were defined through the computation of breakpoints in regression relationships using the host LTT plot. The slopes are measured in log10 of the number of new lineages per log10 of million years. The standard error (SE), the degrees of freedom (df = N-2), the F-statistic, adjusted R², and the p-value are given to indicate the significance of the calculated slope.

Node age estimates on different data sets

For the dated trees constructed on our full and reduced data sets, we obtained the tmrca estimates for different clades along the PV tree (Fig. 4 and Table S6). For those nodes for which the posterior distribution of the age was sampled, we show to what extent the distributions overlap in Fig. S14. Overall, the tmrca estimates obtained by using different data sets are similar. Larger differences are found for the more ancient nodes: the root of the tree, *AvesTestudines/Mammals*, *Aves/Testudines*, *Aves*, and *Mammals* (Fig. 4 and Fig. S14). For these nodes, we obtained older tmrca estimates when using the reduced data sets (RD1 and RD2), as compared to the full data sets (FD and FDF). Due to the scarce number of PVs described for the hosts corresponding to the underlying PV clades, the representative extant taxa are the same in these data sets (see Table S1). Therefore, the

ages of these ancient nodes appear to be mostly influenced by the number of mammalian PVs present in the data sets, where removal of overrepresented PVs leads to older time estimates. For the youngest nodes, *Homo/Colobus*, *Homo/Pan* and *Pan troglodytes/Pan paniscus*, inconsistencies in the tmrca estimates were also obtained (Fig. 4 and Fig. S14). For *Homo/Colobus* and *Pan troglodytes/Pan paniscus*, we obtained younger tmrca estimates when using RD1 and RD2, as compared to the FD and FDF, while for *Homo/Pan* this is the inverse.

Ultimately, we investigated how the TDRexp clock model performs when compared to corrections for time-dependent rate estimates done by fitting a power-law rate decay model to single calibrations under the UCLD model (Supplementary File 1). The overall comparison for the inferred node ages is shown in Fig. 4. For the *Mammals* node (representing the ancestor of PVs infecting mammals) and for most of the underlying nodes, the corrections estimate the node ages to be older than those inferred by the TDRexp model. This is also the case for the *Homo/Colobus*, *Homo/Pan* and *Pan troglodytes/Pan paniscus* nodes. For the more ancient nodes (*Aves/Testudines/Mammals* and *Aves/Testudines*), the corrections estimate the node ages to fall within the range of those inferred by the TDRexp model.

Discussion

With this study we provide an updated overview on the evolutionary history of PVs, with an emphasis in the expansion and radiation events that have punctuated the timeline of this viral family. Our results show that initial PV evolution seems to be marked by virus-host codivergence, where the number of PV lineages slowly increases during the basal diversification of amniotes. The actual increase is probably much faster than the one we communicate here, because a proper sampling of PVs in basal amniotes and in fishes is

426 still wanting. Subsequently, we observe a viral radiation event that appears to be
 427 concomitant with the radiation of the placental mammals. The ecological opportunity for
 428 PVs to explore new hosts (vacant niches), can be considered a trigger of the adaptive
 429 radiation event we observe. Previous studies have already proposed a link between
 430 ecological opportunity and adaptive radiation (Glor, 2010; Yoder et al., 2010). In the case
 431 of PVs, adaptive radiation may increase the chances that some of the descendants are
 432 better able to exploit the new hosts, thus resulting in species that possess different types
 433 of adaptations. When all available niche space becomes filled, we expect rates of lineage
 434 accumulation to decrease (Rabosky & Lovette, 2008), and this is exactly what we observe
 435 for PVs. After the radiation event, the slope of the LTT plot decreases, and in parallel to
 436 virus-host codivergence, independent diversification of PVs occurred. Within the major PV
 437 crown groups, the evolution of key innovations allowed PVs to exploit new resources. At
 438 the molecular level, the most compelling example is that of *AlphaPVs* infecting humans,
 439 where the appearance of the *E5* oncogene (Bravo & Alonso, 2004) triggered an adaptive
 440 radiation that generated three viral lineages with different tissue tropisms. Within these
 441 three clades, PVs further diversified some evolved the potential to degrade tumor
 442 suppressor proteins (Fu et al., 2010; Mesplede et al., 2012; Mirabello et al., 2017; Van
 443 Doorslaer et al., 2015; Willemsen et al., 2019).

444 In this study, we used four different data sets to explore potential biases of the choice of
 445 terminal taxa to infer the evolutionary time scale of PVs. The main differences between the
 446 full and reduced data sets are apparent in (i) the inferred ages of the basal nodes of the
 447 PV tree and (ii) the strength of the time-dependent rate effect. First, the more basal nodes,
 448 between roughly 100 and 550 Ma, are generally inferred to be older when using the
 449 reduced representative sets of taxa compared to using all taxa (Fig. 3b, Fig. 4 and Fig.

450 S14). Secondly, the time-dependent rate effect is stronger when using the reduced data
 451 sets compared to using the full data sets, as evidenced by a steeper slope in Fig. 3a.
 452 Previous studies have reported that phylogenetic error is independent of incomplete taxon
 453 sampling, and that instead, longer sequences will better improve the accuracy of
 454 phylogenetic inference (Rosenberg & Kumar, 2001, 2003). In contrast, other studies have
 455 reported that fewer taxa can lead to increased variance and uncertainty in the results
 456 (Ackerly, 2000), and thus increased taxon sampling is one of the most practical ways to
 457 improve the accuracy of phylogenetic estimates (Heath et al., 2008; Hillis et al., 2003;
 458 Zwickl & Hillis, 2002). With a simulation study where random and non-random samples of
 459 taxa were drawn, it was shown that random sampling provided for statistically robust trait
 460 correlations, whereas non-random sampling (e.g. life-history group or investigator bias) led
 461 to a significant loss of statistical power (Ackerly, 2000). For the construction of the reduced
 462 data sets in this study, it should be noted that taxa pruning was not random, as one
 463 representative was chosen for each PV species and PV type (see Materials and Methods).
 464 Therefore, the same monophyletic lineages of PVs infecting the same hosts were
 465 subsampled for phylogenetic inference with the reduced data sets (RD1 and RD2). RD1
 466 contains representative species with initial shorter branches as compared to those chosen
 467 for RD2. Meaning that for RD2 more diverse PV genomes are being compared, and thus
 468 RD2 represents a more greedy strategy as compared to RD1. Although it has been shown
 469 that greedy algorithms are rarely the best option for statistical power (McAuliffe et al.,
 470 2005), it has also been shown that for genomic comparisons the greedy strategy of
 471 maximizing the evolutionary divergence among species chosen from a known phylogeny
 472 can provide optimal solutions (Pardi & Goldman, 2005). Concordantly, our results indicate
 473 an effect of taxa choice in the reduced data sets, where time inference using single node
 474 calibrations with more diverse terminal taxa (RD2) leads to lower short-term evolutionary

475 rates, than when using less diverse terminal taxa (RD1), while no differences were
 476 observed for the long-term evolutionary rates between both data sets (Fig. 2 and Table
 477 S3). When comparing the results for all data sets in this study, it also seems that RD2
 478 renders closer estimates to those obtained for the full data sets than RD1 does (Fig. 3a-b
 479 and Tables S4-6). In the particular case of this study, increasing the number of taxa
 480 appears to reduce phylogenetic error, leading to more accurate evolutionary rate
 481 estimates, and younger basal node age estimates in the full data sets (FD and FDF). The
 482 combination of multiple data sets in our study also allows us to evaluate the impact on the
 483 estimates for deeper nodes when increasing the number of taxa in the outgroup. In this
 484 case, FD and FDF only differ in the additional presence of four fish PV genomes that were
 485 made available during the study. Indeed, the increased presence of basal fish PVs in the
 486 FDF, results in younger values for the estimate of the age of the root than any other data
 487 set (Fig. 3b). We interpret that the increased number of terminal taxa in the outgroup
 488 decreases the uncertainty about the number and nature of substitutions in this clade
 489 (Heath et al., 2008), leading to a better estimate of the age of the root.

490 Previous studies have shown that different substitution rates can be obtained for the same
 491 species depending on whether they are inferred on a long-term or short-term evolutionary
 492 scale. Here we demonstrate that such time-dependency of inferred evolutionary rates is
 493 also recovered for PVs. The obtained long-term evolutionary rate estimates are in line with
 494 previous overall rates obtained for the *E1* (7.1×10^{-9}) and *L1* (9.6×10^{-9}) genes of PVs (Shah
 495 et al., 2010). Other PV rates estimates, such as 1.84×10^{-8} for recent evolution in the
 496 HPV16 lineage (Pimenoff et al., 2017) and 1.95×10^{-8} for the evolution of a monophyletic
 497 assembly of PVs infecting felids (Rector et al., 2007), fall between the long-term and short-
 498 term rates obtained in this study. This difference can be explained by the inference of
 499 evolutionary rates in more recent times in these studies (for the HPV16 clade, using a data

500 set spanning 0.7 Ma, and for the feline PV clade, using a data set spanning 12 Ma,
501 respectively). Overall, the interpretation of the contrasting estimates in the literature under
502 the framework presented here –a time-dependent behaviour for the inferred evolutionary
503 rates for PVs– supports the hypothesis that a time dependency of the evolutionary rate is
504 present within PVs.

505 With our analyses in [Supplementary File 1](#), we show that the time-dependent rate
506 phenomenon can be well described by fitting a power-law model to the observed data
507 when applying individual calibration nodes. However, the increase in available statistical
508 models and inference tools for complex evolutionary processes allowed us to incorporate
509 the TDR model for time inference of PVs. In our analyses the TDR model performed
510 considerably better than other models that do not allow for rate variation over time. We
511 compared a uniformly and exponentially distributed TDR model. Even though the uniformly
512 distributed model contained more and better dispersed transition times, we find better
513 model fit for the exponential model. This difference in model fit may therefore imply that the
514 rate variation through time may not be as regular as expected from a power-law function
515 (Membrebe et al., 2019). Nevertheless, the increased complexity in the uniform TDR
516 model (*i.e.* more transition times as compared to the exponential TDR model), can also
517 lead to a worse fit of this model.

518 Regarding the evolution of substitution rates, it does not seem that PV evolution has
519 reached an evolutionary stasis, as the PV substitution rates still vary after more than 400
520 million years of virus-host co-evolution ([Supplementary File 1](#)). Episodic and rare events
521 eventually triggering an elevated substitution rate such as host jumps, switch of tissue
522 tropism (Van Doorslaer et al., 2015; Willemsen & Bravo, 2019) and gene gain/loss events
523 (Van Doorslaer & McBride, 2016; Willemsen et al., 2019) might prevent PVs from reaching

524 an evolutionary stasis with their hosts, thus maintaining recent increased substitution
525 rates.

526 We would like to point out that our study suffers from a number of limitations. Our
527 phylogenetic analyses violate the assumptions of extinction (by choosing the Yule tree
528 prior) and complete taxon sampling. In the full data sets we have only worked with
529 reference viral genomes and have not considered the within-type viral diversity. Our
530 jackknifing approach for the reduced data sets contributes itself to this incomplete
531 sampling because we have explicitly chosen to not include all viral taxa. It is therefore that
532 we tested the Yule tree prior versus the Bayesian Skyline tree prior, and the simpler Yule
533 model rendered the best fit to the data. Consequently, the LTT plots in this study are a
534 crude approximation to the speciation rate. Moreover, we used the host LTT plot for
535 identifying changes in the speciation rates over time. However, the precise dating of
536 mammalian evolution as well as the time line of speciation events are a matter of debate
537 for specialists in the field (e.g. Springer et al., 2019). Therefore, our vertebrate taxa choice
538 is biased towards species known to be the hosts of PVs for which the genome has been
539 sequenced, and the breakpoints inferred on the host LTT plot cannot be automatically
540 extended to match and describe the full vertebrate, amniote nor mammalian evolution.
541 Under these limitations, the analysis of the slope time trend in the LTT plots risks of not
542 being powerful enough and to suffer from false negative results. Notwithstanding, we were
543 are able to identify different periods for the variation of the hosts lineages through time,
544 which are compatible with historically significant changes during the evolution of the hosts,
545 as well as with the observed variation of viral lineages through time.

546 With the current set of available genomes we can distinguish four periods with different
547 diversification rates, along the evolutionary time scale of PVs. Our results suggest that the
548 evolutionary history of PVs is multiphasic, where subsequent radiation events allowed

viruses to adapt to new niches, and a subsequent period where independent diversification between viruses and their hosts occurred within the major virus-host clades. Our results improve and refine the previous evolutionary scenario for PVs, that had suggested a biphasic evolution of PVs, where a primary radiation was directly followed by a secondary diversification event (Ignacio G. Bravo & Felez-Sanchez, 2015). The discovery of novel PVs infecting ancestral hosts, like in the present case fish, allowed for the detection of these events. Further increase in taxon sampling of both mammalian and non-mammalian hosts, together with the implementation of more flexible evolutionary models allowing for burst and decay of speciation rates, will lead to a more detailed overview of the evolutionary history of this successful viral family.

Acknowledgements

The authors declare no conflict of interest. We are grateful to the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) for providing computing and storage resources. The authors acknowledge the IRD itrop HPC (South Green Platform) at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this paper. This work was supported by the European Research Council Consolidator Grant CODOVIREVOL (Contract Number 647916) to IGB and by the European Union Horizon 2020 Marie Skłodowska-Curie research and innovation programme grant ONCOGENEVOL (Contract Number 750180) to AW.

Supplementary Material

Supplementary [Tables S1-S6](#), [Supplementary Figures S1-S14](#), and [Supplementary File 1](#) are available online. The alignments, R scripts, starting trees, and xml files are available at <https://doi.org/10.5281/zenodo.5105951>.

References

- Ackerly, D. D. (2000). Taxon sampling, correlated evolution, and independent contrasts. *Evolution*, 54(5), 1480–1492. <https://doi.org/10.1111/j.0014-3820.2000.tb00694.x>
- Aiewsakun, P., & Katzourakis, A. (2015). Time dependency of foamy virus evolutionary rate estimates. *BMC Evolutionary Biology*, 15(1), 119. <https://doi.org/10.1186/s12862-015-0408-z>
- Aiewsakun, P., & Katzourakis, A. (2016). Time-Dependent Rate Phenomenon in Viruses. *Journal of Virology*, 90(16), 7184–7195. <https://doi.org/10.1128/JVI.00593-16>
- Aiewsakun, P., & Katzourakis, A. (2017). Marine origin of retroviruses in the early Palaeozoic Era. *Nature Communications*, 8(May 2016), 13954. <https://doi.org/10.1038/ncomms13954>
- Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., & Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, 29(9), 2157–2167. <https://doi.org/10.1093/molbev/mss084>
- Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., & Lemey, P. (2013). Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*, 30(2), 239–243. <https://doi.org/10.1093/molbev/mss243>
- Bernard, H.-U., Burk, R. D., Chen, Z., van Doorslaer, K., Hausen, H. Zur, & de Villiers, E.-M. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*, 401(1), 70–79. <https://doi.org/10.1016/j.virol.2010.02.002>
- Biek, R., Pybus, O. G., Lloyd-Smith, J. O., & Didelot, X. (2015). Measurably evolving pathogens in the genomic era. In *Trends in Ecology and Evolution* (Vol. 30, Issue 6, pp. 306–313). Elsevier Ltd. <https://doi.org/10.1016/j.tree.2015.03.009>
- Bravo, I. G., & Alonso, A. (2004). Mucosal Human Papillomaviruses Encode Four Different E5 Proteins Whose Chemistry and Phylogeny Correlate with Malignant or Benign Growth. *Journal of Virology*, 78(24), 13613–13626. <https://doi.org/10.1128/JVI.78.24.13613-13626.2004>
- Bravo, Ignacio G., & Felez-Sanchez, M. (2015). Papillomaviruses: Viral evolution, cancer and evolutionary medicine. *Evolution, Medicine and Public Health*, 2015(1), 32–51. <https://doi.org/10.1093/emph/eov003>
- Burk, R. D., Harari, A., & Chen, Z. (2013). Human papillomavirus genome variants. *Virology*, 445(1–2), 232–243. <https://doi.org/10.1016/j.virol.2013.07.018>

607 BurrIDGE, C. P., CRAW, D., FLETCHER, D., & WATERS, J. M. (2008). Geological dates and
608 molecular rates: Fish DNA sheds light on time dependency. *Molecular Biology and*
609 *Evolution*, 25(4), 624–633. <https://doi.org/10.1093/molbev/msm271>

610 BZHALAVA, D., EKLUND, C., & DILLNER, J. (2015). International standardization and
611 classification of human papillomavirus types. *Virology*, 476, 341–344.
612 <https://doi.org/10.1016/j.virol.2014.12.028>

613 CAMPO, M. S. (1997). Bovine papillomavirus and cancer. *The Veterinary Journal*, 154(3),
614 175–188. [https://doi.org/10.1016/S1090-0233\(97\)80019-6](https://doi.org/10.1016/S1090-0233(97)80019-6)

615 CASTRESANA, J. (2000). Selection of Conserved Blocks from Multiple Alignments for Their
616 Use in Phylogenetic Analysis. *Molecular Biology and Evolution*, 17(4), 540–552.
617 <https://doi.org/10.1093/oxfordjournals.molbev.a026334>

618 COMAS, I., COSCOLLA, M., LUO, T., BORRELL, S., HOLT, K. E., KATO-MAEDA, M., PARKHILL, J., MALLA,
619 B., BERG, S., THWAITES, G., YEBOAH-MANU, D., BOTHAMLEY, G., MEI, J., WEI, L., BENTLEY,
620 S., HARRIS, S. R., NIEMANN, S., DIEL, R., ASEFFA, A., ... GAGNEUX, S. (2013). Out-of-
621 Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with
622 modern humans. *Nature Genetics*, 45(10), 1176–1182.
623 <https://doi.org/10.1038/ng.2744>

624 DARRIBA, D., TABOADA, G. L., DOALLO, R., & POSADA, D. (2011). ProtTest 3: fast selection of
625 best-fit models of protein evolution. *Bioinformatics*, 27(8), 1164–1165.
626 <https://doi.org/10.1093/bioinformatics/btr088>

627 DE VILLIERS, E. M., FAUQUET, C., BROKER, T. R., BERNARD, H. U., & ZUR HAUSEN, H. (2004).
628 Classification of papillomaviruses. *Virology*, 324(1), 17–27.
629 <https://doi.org/10.1016/j.virol.2004.03.033>

630 DRUMMOND, A. J., RAMBAUT, A., SHAPIRO, B., & PYBUS, O. G. (2005). Bayesian coalescent
631 inference of past population dynamics from molecular sequences. *Molecular Biology*
632 *and Evolution*, 22(5), 1185–1192. <https://doi.org/10.1093/molbev/msi103>

633 DRUMMOND, ALEXEI J, HO, S. Y. W., PHILLIPS, M. J., & RAMBAUT, A. (2006). Relaxed
634 phylogenetics and dating with confidence. *PLoS Biology*, 4(5), 699–710.
635 <https://doi.org/10.1371/journal.pbio.0040088>

636 DRUMMOND, ALEXEI J, SUCHARD, M. A., XIE, D., & RAMBAUT, A. (2012). Bayesian
637 Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*,
638 29(8), 1969–1973. <https://doi.org/10.1093/molbev/mss075>

639 DUCHENE, S., HOLMES, E. C., & HO, S. Y. W. (2014). Analyses of evolutionary dynamics in
640 viruses are hindered by a time-dependent bias in rate estimates. *Proceedings of the*
641 *Royal Society B: Biological Sciences*, 281(1786), 20140732–20140732.
642 <https://doi.org/10.1098/rspb.2014.0732>

643 Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in
644 viruses: patterns and determinants. *Nature Reviews Genetics*, 9(4), 267–276.
645 <https://doi.org/10.1038/nrg2323>

646 Féléz-Sánchez, M., Trösemeier, J.-H., Bedhomme, S., González-Bravo, M. I., Kamp, C., &
647 Bravo, I. G. (2015). Cancer, Warts, or Asymptomatic Infections: Clinical Presentation
648 Matches Codon Usage Preferences in Human Papillomaviruses. *Genome Biology
649 and Evolution*, 7(8), 2117–2135. <https://doi.org/10.1093/gbe/evv129>

650 Feng, L., Reeves, P. R., Lan, R., Ren, Y., Gao, C., Zhou, Z., Ren, Y., Cheng, J., Wang, W.,
651 Wang, J., Qian, W., Li, D., & Wang, L. (2008). A Recalibrated Molecular Clock and
652 Independent Origins for the Cholera Pandemic Clones. *PLoS ONE*, 3(12), e4053.
653 <https://doi.org/10.1371/journal.pone.0004053>

654 Fu, L., Van Doorslaer, K., Chen, Z., Ristriani, T., Masson, M., Travé, G., & Burk, R. D.
655 (2010). Degradation of p53 by Human Alphapapillomavirus E6 Proteins Shows a
656 Stronger Correlation with Phylogeny than Oncogenicity. *PLoS ONE*, 5(9), e12816.
657 <https://doi.org/10.1371/journal.pone.0012816>

658 García-Moreno, J. (2004). Is there a universal mtDNA clock for birds? *Journal of Avian
659 Biology*, 35(6), 465–468. <https://doi.org/10.1111/j.0908-8857.2004.03316.x>

660 García-Vallvé, S., Alonso, Á., & Bravo, I. G. (2005). Papillomaviruses: different genes have
661 different histories. *Trends in Microbiology*, 13(11), 514–521.
662 <https://doi.org/10.1016/j.tim.2005.09.003>

663 Gernhard, T. (2008). The conditioned reconstructed process. *Journal of Theoretical
664 Biology*, 253(4), 769–778. <https://doi.org/10.1016/j.jtbi.2008.04.005>

665 Gibbs, A. J., Fargette, D., García-Arenal, F., & Gibbs, M. J. (2010). Time - The emerging
666 dimension of plant virus studies. In *Journal of General Virology* (Vol. 91, Issue 1, pp.
667 13–22). <https://doi.org/10.1099/vir.0.015925-0>

668 Glor, R. E. (2010). Phylogenetic Insights on Adaptive Radiation. *Annual Review of
669 Ecology, Evolution, and Systematics*, 41(1), 251–270.
670 <https://doi.org/10.1146/annurev.ecolsys.39.110707.173447>

671 Gottschling, M., Goker, M., Stamatakis, A., Bininda-Emonds, O. R. P., Nindl, I., & Bravo, I.
672 G. (2011). Quantifying the Phylodynamic Forces Driving Papillomavirus Evolution.
673 *Molecular Biology and Evolution*, 28(7), 2101–2113.
674 <https://doi.org/10.1093/molbev/msr030>

675 Gottschling, M., Stamatakis, A., Nindl, I., Stockfleth, E., Alonso, A., & Bravo, I. G. (2007).
676 Multiple Evolutionary Mechanisms Drive Papillomavirus Diversification. *Molecular
677 Biology and Evolution*, 24(5), 1242–1258. <https://doi.org/10.1093/molbev/msm039>

678 Gottschling, Marc, Bravo, I. G., Schulz, E., Bracho, M. A., Deaville, R., Jepson, P. D.,
679 Bresse, M. F. Van, Stockfleth, E., & Nindl, I. (2011). Modular organizations of novel

680 cetacean papillomaviruses. *Molecular Phylogenetics and Evolution*, 59(1), 34–42.
681 <https://doi.org/10.1016/j.ympev.2010.12.013>

682 Heath, T. A., Hedtke, S. M., & Hillis, D. M. (2008). Taxon sampling and the accuracy of
683 phylogenetic analyses. *Journal of Systematics and Evolution*, 46(3), 239–257.
684 <https://doi.org/10.3724/SP.J.1002.2008.08016>

685 Hillis, D. M., Pollock, D. D., McGuire, J. A., & Zwickl, D. J. (2003). Is sparse taxon
686 sampling a problem for phylogenetic inference? In *Systematic Biology* (Vol. 52, Issue
687 1, pp. 124–126). Taylor and Francis Inc. <https://doi.org/10.1080/10635150390132911>

688 Ho, S. Y. W., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., &
689 Cooper, A. (2011). Time-dependent rates of molecular evolution. *Molecular Ecology*,
690 20(15), 3087–3101. <https://doi.org/10.1111/j.1365-294X.2011.05178.x>

691 Ho, S. Y. W., Phillips, M. J., Cooper, A., & Drummond, A. J. (2005). Time dependency of
692 molecular rate estimates and systematic overestimation of recent divergence times.
693 *Molecular Biology and Evolution*, 22(7), 1561–1568.
694 <https://doi.org/10.1093/molbev/msi145>

695 Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software
696 Version 7: Improvements in Performance and Usability. *Molecular Biology and*
697 *Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>

698 Kreider, J. W., & Bartlett, G. L. (1981). The Shope papilloma-carcinoma complex of rabbits:
699 A model system of neoplastic progression and spontaneous regression. *Advances in*
700 *Cancer Research*, 35(C), 81–110. [https://doi.org/10.1016/S0065-230X\(08\)60909-4](https://doi.org/10.1016/S0065-230X(08)60909-4)

701 Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for
702 Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, 34(7),
703 1812–1819. <https://doi.org/10.1093/molbev/msx116>

704 Li, W. L. S., & Drummond, A. J. (2012). Model Averaging and Bayes Factor Calculation of
705 Relaxed Molecular Clocks in Bayesian Phylogenetics. *Molecular Biology and*
706 *Evolution*, 29(2), 751–761. <https://doi.org/10.1093/molbev/msr232>

707 López-Bueno, A., Mavian, C., Labella, A. M., Castro, D., Borrego, J. J., Alcami, A., & Alejo,
708 A. (2016). Concurrence of Iridovirus, Polyomavirus, and a Unique Member of a New
709 Group of Fish Papillomaviruses in Lymphocystis Disease-Affected Gilthead Sea
710 Bream. *Journal of Virology*, 90(19), 8768–8779. <https://doi.org/10.1128/JVI.01369-16>

711 McAuliffe, J. D., Jordan, M. I., & Pachter, L. (2005). Subtree power analysis and species
712 selection for comparative genomics. *Proceedings of the National Academy of*
713 *Sciences*, 102(22), 7900–7905. <https://doi.org/10.1073/pnas.0502790102>

714 Membrebe, J. V., Suchard, M. A., Rambaut, A., Baele, G., & Lemey, P. (2019). Bayesian
715 Inference of Evolutionary Histories under Time-Dependent Substitution Rates.

716 *Molecular Biology and Evolution*, 36(8), 1793–1803.
717 <https://doi.org/10.1093/molbev/msz094>

718 Mengual-Chuliá, B., Domenis, L., Robetto, S., & Bravo, I. G. (2014). A novel papillomavirus
719 isolated from a nasal neoplasia in an Italian free-ranging chamois (*Rupicapra r.*
720 *rupicapra*). *Veterinary Microbiology*, 172(1–2), 108–119.
721 <https://doi.org/10.1016/j.vetmic.2014.05.006>

722 Mesplede, T., Gagnon, D., Bergeron-Labrecque, F., Azar, I., Senechal, H., Coutlee, F., &
723 Archambault, J. (2012). p53 Degradation Activity, Expression, and Subcellular
724 Localization of E6 Proteins from 29 Human Papillomavirus Genotypes. *Journal of*
725 *Virology*, 86(1), 94–107. <https://doi.org/10.1128/JVI.00751-11>

726 Mirabello, L., Yeager, M., Yu, K., Clifford, G. M., Xiao, Y., Zhu, B., Cullen, M., Boland, J. F.,
727 Wentzensen, N., Nelson, C. W., Raine-Bennett, T., Chen, Z., Bass, S., Song, L., Yang,
728 Q., Steinberg, M., Burdett, L., Dean, M., Roberson, D., ... Schiffman, M. (2017).
729 HPV16 E7 Genetic Conservation Is Critical to Carcinogenesis. *Cell*, 170(6), 1164-
730 1174.e6. <https://doi.org/10.1016/j.cell.2017.08.001>

731 Monographs, I. (2012). Biological Agents, Human Papillomaviruses. *IARC Monographs on*
732 *the Evaluation of Carcinogenic Risks to Humans*, 100B. [https://monographs.iarc.fr/wp-](https://monographs.iarc.fr/wp-content/uploads/2018/06/mono100B-11.pdf)
733 [content/uploads/2018/06/mono100B-11.pdf](https://monographs.iarc.fr/wp-content/uploads/2018/06/mono100B-11.pdf)

734 Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial
735 molecular clock: The mid-aegean trench calibration. *Molecular Biology and Evolution*,
736 27(7), 1659–1672. <https://doi.org/10.1093/molbev/msq051>

737 Pardi, F., & Goldman, N. (2005). Species Choice for Comparative Genomics: Being
738 Greedy Works. *PLoS Genetics*, 1(6), e71.
739 <https://doi.org/10.1371/journal.pgen.0010071>

740 Pattengale, N. D., Aberer, A. J., Swenson, K. M., Stamatakis, A., & Moret, B. M. E. (2011).
741 Uncovering Hidden Phylogenetic Consensus in Large Data Sets. *IEEE/ACM*
742 *Transactions on Computational Biology and Bioinformatics*, 8(4), 902–911.
743 <https://doi.org/10.1109/TCBB.2011.28>

744 Pimenoff, V. N., de Oliveira, C. M., & Bravo, I. G. (2017). Transmission between Archaic
745 and Modern Human Ancestors during the Evolution of the Oncogenic Human
746 Papillomavirus 16. *Molecular Biology and Evolution*, 34(1), 4–19.
747 <https://doi.org/10.1093/molbev/msw214>

748 Puigbo, P., Garcia-Vallve, S., & McInerney, J. O. (2007). TOPD/FMTS: a new software to
749 compare phylogenetic trees. *Bioinformatics*, 23(12), 1556–1558.
750 <https://doi.org/10.1093/bioinformatics/btm135>

751 R Core Team. (2014). *R: A language and environment for statistical computing*. R
752 Foundation for Statistical Computing. <http://www.r-project.org/>

- 753 Rabosky, D. L., & Lovette, I. J. (2008). Density-dependent diversification in North American
754 wood warblers. *Proceedings of the Royal Society B: Biological Sciences*, 275(1649),
755 2363–2371. <https://doi.org/10.1098/rspb.2008.0630>
- 756 Rector, A., Lemey, P., Tachezy, R., Mostmans, S., Ghim, S.-J., Van Doorslaer, K., Roelke,
757 M., Bush, M., Montali, R. J., Joslin, J., Burk, R. D., Jenson, A. B., Sundberg, J. P.,
758 Shapiro, B., & Van Ranst, M. (2007). Ancient papillomavirus-host co-speciation in
759 Felidae. *Genome Biology*, 8(4), R57. <https://doi.org/10.1186/gb-2007-8-4-r57>
- 760 Rector, A., Stevens, H., Lacave, G., Lemey, P., Mostmans, S., Salbany, A., Vos, M., Van
761 Doorslaer, K., Ghim, S.-J., Rehtanz, M., Bossart, G. D., Jenson, A. B., & Van Ranst,
762 M. (2008). Genomic characterization of novel dolphin papillomaviruses provides
763 indications for recombination within the Papillomaviridae. *Virology*, 378(1), 151–161.
764 <https://doi.org/10.1016/j.virol.2008.05.020>
- 765 Robles-Sikisaka, R., Rivera, R., Nollens, H. H., St. Leger, J., Durden, W. N., Stolen, M.,
766 Burchell, J., & Wellehan, J. F. X. (2012). Evidence of recombination and positive
767 selection in cetacean papillomaviruses. *Virology*, 427(2), 189–197.
768 <https://doi.org/10.1016/j.virol.2012.01.039>
- 769 Rocha, E. P. C., Smith, J. M., Hurst, L. D., Holden, M. T. G., Cooper, J. E., Smith, N. H., &
770 Feil, E. J. (2006). Comparisons of dN/dS are time dependent for closely related
771 bacterial genomes. *Journal of Theoretical Biology*, 239(2), 226–235.
772 <https://doi.org/10.1016/j.jtbi.2005.08.037>
- 773 Rosenberg, M. S., & Kumar, S. (2001). Incomplete taxon sampling is not a problem for
774 phylogenetic inference. *Proceedings of the National Academy of Sciences of the*
775 *United States of America*, 98(19), 10751–10756.
776 <https://doi.org/10.1073/pnas.191248498>
- 777 Rosenberg, M. S., & Kumar, S. (2003). Taxon sampling, bioinformatics, and
778 phylogenomics. In *Systematic Biology* (Vol. 52, Issue 1, pp. 119–124). Taylor and
779 Francis Inc. <https://doi.org/10.1080/10635150390132894>
- 780 Scase, T., Brandt, S., Kainzbauer, C., Sykora, S., Bijmolt, S., Hughes, K., Sharpe, S., &
781 Foote, A. (2010). Equus caballus papillomavirus-2 (EcPV-2): an infectious cause for
782 equine genital cancer? *Equine Veterinary Journal*, 42(8), 738–745.
783 <https://doi.org/10.1111/j.2042-3306.2010.00311.x>
- 784 Shah, S. D., Doorbar, J., & Goldstein, R. A. (2010). Analysis of host-parasite incongruence
785 in papillomavirus evolution using importance sampling. *Molecular Biology and*
786 *Evolution*, 27(6), 1301–1314. <https://doi.org/10.1093/molbev/msq015>
- 787 Shimodaira, H., & Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with
788 Applications to Phylogenetic Inference. *Molecular Biology and Evolution*, 16(8), 1114–
789 1116. <https://doi.org/10.1093/oxfordjournals.molbev.a026201>

790 Simmonds, P., Aiewsakun, P., & Katzourakis, A. (2019). Prisoners of war — host
791 adaptation and its constraints on virus evolution. *Nature Reviews Microbiology*, 17(5),
792 321–328. <https://doi.org/10.1038/s41579-018-0120-2>

793 Springer, M. S., Foley, N. M., Brady, P. L., Gatesy, J., & Murphy, W. J. (2019). Evolutionary
794 Models for the Diversification of Placental Mammals Across the KPg Boundary. In
795 *Frontiers in Genetics* (Vol. 10, p. 1241). Frontiers Media S.A.
796 <https://doi.org/10.3389/fgene.2019.01241>

797 Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis
798 of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
799 <https://doi.org/10.1093/bioinformatics/btu033>

800 Subramanian, S., Denver, D. R., Millar, C. D., Heupink, T., Aschrafi, A., Emslie, S. D.,
801 Baroni, C., & Lambert, D. M. (2009). High mitogenomic evolutionary rates and time
802 dependency. In *Trends in Genetics* (Vol. 25, Issue 11, pp. 482–486).
803 <https://doi.org/10.1016/j.tig.2009.09.005>

804 Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A.
805 (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10.
806 *Virus Evolution*, 4(1). <https://doi.org/10.1093/ve/vey016>

807 Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein
808 sequence alignments into the corresponding codon alignments. *Nucleic Acids*
809 *Research*, 34(Web Server), W609–W612. <https://doi.org/10.1093/nar/gkl315>

810 Tisza, M. J., Pastrana, D. V., Welch, N. L., Stewart, B., Peretti, A., Starrett, G. J., Pang, Y.
811 Y. S., Krishnamurthy, S. R., Pesavento, P. A., McDermott, D. H., Murphy, P. M.,
812 Whited, J. L., Miller, B., Brenchley, J., Rosshart, S. P., Rehmann, B., Doorbar, J.,
813 Ta'ala, B. A., Pletnikova, O., ... Buck, C. B. (2020). Discovery of several thousand
814 highly diverse circular DNA viruses. *ELife*, 9. <https://doi.org/10.7554/eLife.51971>

815 Van Doorslaer, K., Chen, Z., Bernard, H. U., Chan, P. K. S., Desalle, R., Dillner, J.,
816 Forslund, O., Haga, T., McBride, A. A., Villa, L. L., & Burk, R. D. (2018). ICTV virus
817 taxonomy profile: Papillomaviridae. *Journal of General Virology*, 99(8), 989–990.
818 <https://doi.org/10.1099/jgv.0.001105>

819 Van Doorslaer, K., DeSalle, R., Einstein, M. H., & Burk, R. D. (2015). Degradation of
820 Human PDZ-Proteins by Human Alphapapillomaviruses Represents an Evolutionary
821 Adaptation to a Novel Cellular Niche. *PLOS Pathogens*, 11(6), e1004980.
822 <https://doi.org/10.1371/journal.ppat.1004980>

823 Van Doorslaer, K., Li, Z., Xirasagar, S., Maes, P., Kaminsky, D., Liou, D., Sun, Q., Kaur, R.,
824 Huyen, Y., & McBride, A. A. (2017). The Papillomavirus Episteme: A major update to
825 the papillomavirus sequence database. *Nucleic Acids Research*, 45(D1), D499–D506.
826 <https://doi.org/10.1093/nar/gkw879>

827 Van Doorslaer, K., & McBride, A. A. (2016). Molecular archeological evidence in support of
828 the repeated loss of a papillomavirus gene. *Scientific Reports*, 6(1), 33028.
829 <https://doi.org/10.1038/srep33028>

830 Van Doorslaer, K., Ruoppolo, V., Schmidt, A., Lescroël, A., Jongsomjit, D., Elrod, M.,
831 Kraberger, S., Stainton, D., Dugger, K. M., Ballard, G., Ainley, D. G., & Varsani, A.
832 (2017). Unique genome organization of non-mammalian papillomaviruses provides
833 insights into the evolution of viral early proteins. *Virus Evolution*, 3(2), vex027.
834 <https://doi.org/10.1093/ve/vex027>

835 Willemsen, A., & Bravo, I. G. (2019). Origin and evolution of papillomavirus (onco)genes
836 and genomes. *Philosophical Transactions of the Royal Society of London. Series B,*
837 *Biological Sciences*, 374(1773), 20180303. <https://doi.org/10.1098/rstb.2018.0303>

838 Willemsen, A., Félez-Sánchez, M., & Bravo, I. (2019). Genome plasticity in
839 Papillomaviruses and de novo emergence of E5 oncogenes. *Genome Biology and*
840 *Evolution*, *accepted*, evz095. <https://doi.org/10.1093/gbe/evz095>

841 Yoder, J. B., Clancey, E., Des Roches, S., Eastman, J. M., Gentry, L., Godsoe, W., Hagey,
842 T. J., Jochimsen, D., Oswald, B. P., Robertson, J., Sarver, B. A. J., Schenk, J. J.,
843 Spear, S. F., & Harmon, L. J. (2010). Ecological opportunity and the origin of adaptive
844 radiations. *Journal of Evolutionary Biology*, 23(8), 1581–1596.
845 <https://doi.org/10.1111/j.1420-9101.2010.02029.x>

846 Yule, G. U. (1925). A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J.
847 C. Willis, F.R.S. *Philosophical Transactions of the Royal Society B: Biological*
848 *Sciences*, 213(402–410), 21–87. <https://doi.org/10.1098/rstb.1925.0002>

849 Zwickl, D. J., & Hillis, D. M. (2002). Increased taxon sampling greatly reduces phylogenetic
850 error. *Systematic Biology*, 51(4), 588–598. <https://doi.org/10.1080/10635150290102339>
851

852 **Figure Legends**

853 **Figure 1.** Dated Bayesian phylogenetic tree for reduced data set 2 (RD2) containing 130
 854 PVs. The tree was constructed at the nucleotide level based on the concatenated *E1-E2-*
 855 *L2-L1* genes, using an exponential TDR model and a Yule tree prior. The scale bar is given
 856 in million years ago (Ma). Values at the nodes correspond to posterior probabilities, where
 857 asterisks indicate full support. Error bars encompass 95% highest posterior density (HPD)
 858 intervals for the age of the nodes. Clock symbols indicate the nodes used for calibration.
 859 The clades are coloured according to the PV crown group classification, as indicated in the
 860 legend on the left. Next to the tree on the right, the taxonomic host group (superorder,
 861 superclass, class, order, parvorder, no rank) corresponds to the one in which the
 862 corresponding host clades could best be summarized. The cartoon silhouettes illustrate
 863 some distinctive members of these clades. The grey shaded areas (delimited by vertical
 864 black dashed lines) give an overall overview of four periods that can be distinguished
 865 along the evolutionary timeline of PVs, as shown in [Fig. 3c](#).

866 **Figure 2.** Time-dependent rate phenomenon for PVs. **a)** Results are given for 18
 867 phylogenetic time inferences on RD1. Each inference was performed with a single
 868 calibration node, for which the numbers in the legend correspond to those in [Table 1](#). The
 869 x-axis represents time from past to present in million years ago (Ma). The y-axis
 870 represents the evolutionary rate in substitutions/site/million years (s/s/my). Error bars
 871 correspond to the mean and standard deviation of both the inferred node time and the
 872 inferred substitution rate. The black line displays the trend of a linear model fit to the data,
 873 with the grey shaded area indicating the 95% confidence interval. **b)** The same analysis as
 874 in panel a was performed on RD2. Note that the y-axes in panels a and b are not on the
 875 same scale.

Figure 3. Time-dependent rate epoch modelling and diversification on dated phylogenetic PV trees. Bayesian time inference was performed using two representative reduced data sets (RD1 and RD2) and two full data sets (FD and FDF). **a)** Time-dependent rate effect under an epoch structure with exponentially distributed time intervals. The x-axis represents time from past to present in million years ago (Ma). The y-axis represents the evolutionary rate in substitutions/site/million years (s/s/my). The error bars represent the 95% highest posterior density (HPD) intervals of the rate estimates. Both axes are on a log₁₀ scale. **b)** Dated phylogenetic trees for each of the datasets used, inferred with an exponential TDR model and a Yule tree prior. The x-axis represents time from past to present in Ma. The arrows indicate three basal nodes for which the inferred age is given with the 95% HPD between brackets. These basal nodes are inferred to be much older in RD1 and RD2 when compared to FD and FDF. **c)** Lineages through time (LTT) plot for RD1, RD2, FD, FDF and the host tree. The x-axis represents time from past to present in Ma. The y-axis represents the number of lineages. Both axes are on a log₁₀ scale. The grey vertical bars and black dashed vertical lines separate four evolutionary periods (I-IV) that we observe in the viral and in the host data, respectively. **d)** The slopes of the LTT lines were calculated using the four evolutionary periods of the host data. The slopes are also provided in Table 4. This figure highlights the differences in results obtained for the different data sets, but more importantly, shows that PVs have not diversified at a constant rate.

Figure 4. Inferred age for ancestral nodes on dated Bayesian phylogenetic trees. This figure allows to compare for each node the inferred age on the different data sets: RD1, RD2, FD, and FDF. Dated phylogenetic trees were inferred by using the exponential TDR

899 model with a Yule tree prior and 18 calibration nodes. RD1_corr and RD2_corr are
900 corrections for the time-dependent rate phenomenon for trees inferred by time inference
901 using an UCLD model with a Yule tree prior and 18 single calibration nodes
902 ([Supplementary File 1](#)). The nodes are ancestors of clades within the PV crown groups and
903 unclassified clades, matched in colour code with [Fig. 1](#). The nodes used for calibration are
904 indicated with a clock symbol and the corresponding number (see [Table 1](#)). The node age
905 is in million years ago (Ma) and on a log10 scale. For RD1, RD2, FD and FDF, error bars
906 encompass 95% highest posterior density (HPD) intervals for the age of the nodes. For
907 RD1_corr and RD2_corr, error bars encompass the lowest and highest inferred 95%
908 confidence intervals (CI) for the 18 different nodes used for error correction. The asterisk
909 indicates that not all the extant PV lineages underlying this node infect mammals as there
910 is one exception in FD and FDF data sets: a PV isolated from a carpet python (MsPV1).

Figure 1

Figure 2**a)**

RD1

b)

RD2

Figure 3**a)** **b)** **c)** **d)**

Figure 4