bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Co-evolutionary analysis suggests a role for TLR9 in papillomavirus restriction

Kelly King[1], Brendan B. Larsen[2], Sophie Gryseels[2,3,4], Cécile Richet[5], Simona Kraberger[5], Robert Jackson[1], Michael Worobey[2,6], Joseph S. Harrison[7], Arvind Varsani[5,8], Koenraad Van Doorslaer[1,6,9]*

[1]School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721, USA

[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA.

[3]Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, 3000 Leuven, Belgium.

[4]Department of Biology, University of Antwerp, 2000 Antwerp, Belgium.

[5]The Biodesign Center for Fundamental and Applied Microbiomics, Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287-5001, USA

[6]The BIO5 Institute, University of Arizona, Tucson, Arizona 85721, USA.

[7]Department of Chemistry, University of the Pacific, Stockton, California, USA.

[8]Structural Biology Research Unit, Department of Integrative Biomedical Sciences, University of Cape Town, Observatory, Cape Town 7701, South Africa

[9]Department of Immunobiology; Cancer Biology Graduate Interdisciplinary Program; UA Cancer Center, University of Arizona Tucson, AZ 85724, USA

* Correspondence: vandoorslaer@arizona.edu

Keywords: *Papillomaviridae*; Yinpterochiroptera; Yangochiroptera; Mexican free-tailed bat; innate immunity, TLR9

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A. Abstract

Upon infection, DNA viruses can be sensed by pattern recognition receptors (PRRs) leading to the activation of type I and III interferons, aimed at blocking infection. Therefore, viruses must inhibit these signaling pathways, avoid being detected, or both. Papillomavirus virions are trafficked from early endosomes to the Golgi apparatus and wait for the onset of mitosis to complete nuclear entry. This unique subcellular trafficking strategy avoids detection by cytoplasmic PRRs, a property that may contribute to establishment of infection. However, as the capsid uncoats within acidic endosomal compartments, the viral DNA may be exposed to detection by toll-like receptor (TLR) 9. In this study we characterize two new papillomaviruses from bats and use molecular archeology to demonstrate that their genomes altered their nucleotide composition to avoid detection by TLR9, providing evidence that TLR9 acts as a PRR during papillomavirus infection. Furthermore, we demonstrate that TLR9, like other components of the innate immune system, is under evolutionary selection in bats, providing the first direct evidence for co-evolution between papillomaviruses and their hosts.

B. Introduction

Papillomaviruses (PVs) are circular double-stranded DNA viruses found in an extensive repertoire of hosts, including mammals, reptiles, birds, and fish (Van Doorslaer 2013; Van Doorslaer et al. 2013; 2017; 2018). In humans, roughly 400 genetically diverse papillomavirus types have been described. While a subset of these viruses is associated with (malignant) tumors, most viral types do not cause disease in immunocompetent hosts. As with humans, hosts that have been thoroughly sampled are infected with an extensive repertoire of highly diverse yet species-specific viruses. Co-evolution of virus and host alone is insufficient to explain the phylogeny of viruses in the family *Papillomaviridae*. For example, papillomaviruses infecting humans do not form a monophyletic group, within the papillomavirus family member phylogenetic tree suggests multiple evolutionary mechanisms associated with host cellular interactions and immune evasion as

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

important factors throughout viral genome evolution (Van Doorslaer 2013; Willemsen and Bravo 2019).

Furthermore, host or tissue tropism is likely a significant determinant of host-pathogen interactions (Carey et al. 2019; Sawyer, Emerman, and Malik 2004; Taubenberger and Kash 2010). Therefore, genomic analyses that consider essential mechanisms of the viral life cycle and evolutionary pressures related to host-parasite interactions may provide a novel perspective into papillomavirus genome evolution. A broader description of animal viruses will continue to inform these efforts.

A successful infection requires that the papillomavirus DNA is delivered to the host cell nucleus. Papillomaviruses access mitotically active basal cells through lesions in the stratified epithelia of cutaneous or mucosal tissues. Following binding to cellular receptors and priming by kallikrein-8 and furin cleavage, the virus is endocytosed (Aksoy, Gottschalk, and Meneses 2017; Day and Schelhaas 2014; DiGiuseppe, Bienkowska-Haba, Guion, and Sapp 2017; Richards et al. 2006; Cerqueira et al. 2015; Day et al. 2013; Schelhaas et al. 2012). The viral DNA is transported to the Golgi before the mitosis-dependent nuclear accumulation of L2 and viral DNA near PML bodies (Day et al. 2013; Lipovsky et al. 2013; Popa et al. 2015; Aydin et al. 2017; 2014; Calton et al. 2017; DiGiuseppe, Bienkowska-Haba, Guion, Keiffer, et al. 2017; Stepp et al. 2017; Day et al. 2004).

Host cells detect a variety of viral pathogen-associated molecular patterns (PAMPs) by pattern-recognition receptors (PRRs), resulting in induction of interferon (IFN) and a potent antiviral response (Medzhitov 2007). The concerted actions of PRR signaling, specific viral-restriction factors, and viral evasion strategies determine the eventual outcome of viral infection (Bowie and Unterholzner 2008).

The papillomaviral structural proteins (L1 and L2) have no known enzymatic activity to directly counteract antiviral responses (Buck, Day, and Trus 2013; J. W. Wang and Roden 2013). Instead, we demonstrated that papillomaviruses evolved an elaborate trafficking mechanism to evade

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

PRR sensing pathways within the cytosol (Uhlorn, Gamez, et al. 2020; Campos 2017; Uhlorn, Jackson, et al. 2020). Furthermore, millions of years of virus-host co-speciation left historical evidence of immune evasion events in these viruses' genomes (Sorouri et al. 2020). For example, APOBEC3 has been demonstrated to restrict infection with HPV (Warren, Xu, et al. 2015; Warren, Van Doorslaer, et al. 2015). We previously demonstrated that alphapapillomaviruses are significantly depleted of TpC dinucleotides, the target for APOBEC3 mediated mutagenesis. This TpC depletion evolved as a mechanism to evade APOBEC3 mediated mutagenesis. Specifically, this depletion of the TpC content is more pronounced in mucosal alphapapillomaviruses and is correlated with significantly higher expression levels of APOBEC3 in mucosal tissues (Warren, Van Doorslaer, et al. 2015). These findings illustrate that host antiviral activity plays a critical role in regulating papillomavirus evolution and that “molecular archeology” can be used to identify these events.

Toll-like receptors (TLR) survey the extracellular and endosomal compartments and represent the first defense line against foreign invaders. TLR2 and TLR4 recognize viral glycoproteins (Blanco et al. 2010; Boehme, Guerrero, and Compton 2006; Jude et al. 2003; Bieback et al. 2002; Murawski et al. 2009; Rassa et al. 2002; M. R. Thompson et al. 2011); TLR3 recognizes double-stranded RNA (Alexopoulou et al. 2001; Bell et al. 2006; Gowen et al. 2006; Oshiumi et al. 2011; F. Weber et al. 2006; Choe 2005), TLR7 and TLR8 recognize viral single-stranded RNA (Akira, Uematsu, and Takeuchi 2006; Diebold 2004; Hemmi et al. 2002; Kawai and Akira 2006; Zucchini et al. 2008; Jurk et al. 2002; Heil 2004). The endosomal TLR9 detects unmethylated CpG motifs found in dsDNA (viral) genomes (Bowie and Unterholzner 2008; Gupta et al. 2015; M. R. Thompson et al. 2011; J. Thompson and Iwasaki 2008).

As described above, HPV particles are endocytosed (Campos 2017; Calton et al. 2017), and viral DNA could be recognized by endosomal TLR9 resulting in a downstream inflammatory immune response. We hypothesize that TLR9 may detect papillomavirus dsDNA leading to CpG depletion, similar to what we observed for APOBEC3 target motifs. Indeed, papillomavirus genomes have

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reduced CpG content (Warren, Van Doorslaer, et al. 2015; Upadhyay and Vivekanandan 2015). This overall dinucleotide depletion confounds the ability to demonstrate the cause of this depletion.

Bats serve as reservoirs for many viruses and and have served as the source of well-documented cross-species transmission events of pathogens responsible for a myriad of epidemics, the most notable being severe acute respiratory syndrome coronavirus 1, middle east respiratory syndrome coronavirus, Ebola virus, Marburg virus, and most recently, severe acute respiratory syndrome coronavirus 2 (Banerjee et al. 2020; Brook and Dobson 2015; Wacharapluesadee et al. 2021). It has been proposed that bats avoid immunopathological outcomes by not fully clearing a viral infection (O'Shea et al. 2014). Indeed, bats have been reported to exhibit a 'dampened' immune response to viral infections (Banerjee et al. 2020; Gorbunova, Seluanov, and Kennedy 2020; Xie et al. 2018; Zhang et al. 2013; Subudhi, Rapin, and Misra 2019). The complex suppression of immune response pathways is variable between several bat species (order Chiroptera) (Zhang et al. 2013; Jiang et al. 2017; Escalera-Zamudio et al. 2015; Hawkins et al. 2019). Importantly, it was demonstrated that residues involved in the ligand-binding region of the bat TLR9 protein are evolving under positive selection (Escalera-Zamudio et al. 2015; Jiang et al. 2017). This evolutionary selection of TLR9 has been proposed to contribute to the high tolerance for viral infections observed in bats (Banerjee et al. 2020; Hawkins et al. 2019). Notably, the theory of co-evolution suggests that viruses need to counter these changes in TLR9, and we should be able to detect these host-parasite interactions (Tan et al. 2017; Warren, Van Doorslaer, et al. 2015).

To address this question, we determined the genomes of two novel bat papillomaviruses from *Tadarida brasiliensis* (TbraPV2, TbraPV3). Taxonomically, bats are classified into two suborders; Yinpterochiroptera and Yangochiroptera. By comparing the genomes of papillomaviruses associated with bats from either suborder, we demonstrate that TLR9 target motifs are significantly depleted and impact papillomavirus evolution. Furthermore, we extend existing data showing that Yangochiroptera TLR9 is evolving under diversifying selection. This argues that

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

papillomavirus genomes are evolving in response to a host change. This is the first direct evidence of PVs evolving in response to host evolutionary changes, thus providing direct evidence for co-evolution. Also, these data argue that TLR9 is a restriction factor for papillomavirus infection.

C. Results

C1. Sampling, sample processing and viral metagenomics

We identified two circular contigs (circular based on terminal redundancy) that had similarities to papillomavirus sequences. We mapped the raw reads to the assembled genomes using BBmap (Bushnell 2014)to determine the read depth. For both the genomes we had a 22-25X coverage depth across the whole genome with 1200-1300 mapped reads.

C2. New bat associated papillomaviruses cluster with previously identified Chiropteran viruses

Using a metagenomics approach, we determined the genomes of two novel circular dsDNA viruses. The open reading frames of these putative new viruses were determined using PuMA (Pace et al., 2020). This analysis identified the typical papillomavirus open reading frames (E6, E7, E1, E2, L1, and L2) and the spliced E1^E4 and E8^E2 mRNAs), suggesting that we recovered the genomes of two papillomaviruses associated with Mexican free-tailed bats (*Tadarida brasiliensis*). The current papillomavirus taxonomy is based on sequence identity across the L1 open reading frame. If two viruses share more than 60%, they fall into the same genus. Species within a genus group viral 'types' that share between 70 and 90% sequence identity. A new papillomavirus type shares less than 90% sequence identity with other viruses (Van Doorslaer et al. 2018; Bernard et al. 2010; de Villiers et al. 2004). Both identified viruses share less than 90% identity with their closest relatives (**Figure 1B**). In consultation with the international animal papillomavirus reference center (Van Doorslaer and Dillner 2019), we name these two novel papillomaviruses TbraPV2 (*GenBank* # MW922427) and TbraPV3 (*GenBank* # MW922428), respectively. TbraPV2 is 8093 bp long, while TbraPV3 is 8037 bp long. Based on pairwise

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sequence identity in the L1 open reading frame, both viruses are most closely related to TbraPV1 (**Figure 1B**). TbraPV2 shares 81.7% sequence identity with TbraPV1 and likely represents a new species in this as of yet unclassified genus. TbraPV3 shares 60.4% identity with TbraPV1, placing it in the same genus. However, the phylogenetic tree shown in **Figure 1** places TbraPV3 as an outgroup to a clade that contains HPV41, EdPV1, TbraPV1, TbraPV2. Therefore, the evolutionary history of these viruses does not support the current L1-based taxonomy.

It has been demonstrated that co-speciation between PVs and their hosts is a major contributor to the papillomavirus's evolutionary history (Van Doorslaer 2013; Gottschling et al. 2007; 2011). In support of this notion, these novel bat papillomaviruses cluster with other previously described Chiropteran papillomaviruses (**Figure 1A**). However, as for other papillomavirus-host relationships, bat-associated viruses are paraphyletic, suggesting that other evolutionary mechanisms like intra-host divergence or niche adaptation likely contribute to the papillomavirus phylogenetic tree (Buck et al. 2016; Van Doorslaer 2013).

C3. Chiropteran PVs co-speciated with their hosts

While TbraPV2 and TbraPV3 cluster together with several other Chiropteran viruses, the larger clade consists of viruses infecting a wide array of mammals (red arrow in **Figure 1A**). In addition to 6 species of Chiroptera, the subtree contains 16 host species classified in 5 mammalian orders. We wanted to compare the evolutionary history of these diverse viruses to their hosts. Due to intra-host divergence and niche adaptation, papillomaviruses infecting the same host can be found in multiple phylogenetic tree clades. To ensure that viruses with a similar tissue tropism and evolutionary history are compared, we extracted a subtree from the maximum likelihood tree (**Figure 2**) (Smeele et al. 2018). This clade contains the newly identified TbraPV2 and TbraPV3 embedded within the largest monophyletic Chiroptera papillomavirus clade (red arrow in **Figure 1A**). We used a tanglegram to address our hypothesis of virus-host co-evolution (**Figure 2A**). In this analysis, nodes in the host and virus phylogeny are rotated to optimize tip matching. Similarities between the host and virus phylogenetic relationships are indicated by parallel lines

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

linking the virus to its host in their respective trees. Conversely, mismatches in the evolutionary history of the host and the virus show overlapping connecting lines. While there are some overlapping connections between papillomaviruses and their hosts, most virus-host pairs support the idea of co-speciation.

To formally quantify the degree of co-speciation, we focused on two datasets. First, we used the phylogenetic tree shown in **Figure 2**. Because it was previously shown that members of the *Lambdapapillomavirus* genus co-evolve with their hosts (Rector et al. 2007), we also tested a smaller subtree to avoid skewing the results (indicated with the red arrow in **Figure 2**). The host and virus phylogenetic trees were compared using the Wasserstein distance, estimated to be 0.205 and 0.284 for the larger and smaller datasets, respectively. A Wasserstein distance of 0 indicates that both trees are topologically identical, while a value of 1 indicates complete lack of congruence between both trees (Lewitus and Morlon 2016). Therefore, the host tree predicts the virus tree, suggesting a role for co-speciation.

Also, we used the Procrustean Approach to Cophylogeny (PACo). This approach evaluates congruency between distance matrices for each virus and associated host phylogenies (Balbuena, Míguez-Lozano, and Blasco-Costa 2013; Hutchinson et al. 2017). The observed best-fit Procrustean super-imposition (1.08E5 and 3.22E4 for the larger and smaller dataset, respectively; denoted by the red dotted line) lies outside of the 95% confidence interval of the ensemble of 1000 network randomizations in the null model (**Figures 2B and C**). Therefore, the data allow us to reject the null hypotheses that the papillomavirus host tree does not predict the virus tree and supports co-speciation as an important process for the evolution of this subclade of PVs and their hosts (Balbuena, Míguez-Lozano, and Blasco-Costa 2013; Hutchinson et al. 2017).

C4. Yangochiropteran viruses have a reduced CpG content.

We previously demonstrated that millions of years of virus-host co-speciation left historical evidence of this virus-host arms-race in the papillomavirus' genomes. For example, the mucosal,

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

cancer-causing alphapapillomaviruses have a reduced TpC dinucleotide content, presumably due to evolutionary adaptations to APOBEC3 editing (Warren, Van Doorslaer, et al. 2015). To extend these studies, we calculated the observed/predicted ratio for each dinucleotide in the viruses shown in **Figure 2**. A ratio close to 1 indicates that a dinucleotide is seen in the sequence as often as expected based on each sequence's nucleotide composition. Values lower than 1 suggest that a dinucleotide is depleted. While the ApC, ApT, GpT, TpA, and TpC ratios are significantly lower than 1 (one-sample t-test p-value < 0.001), we observed the most significant decrease in the CpG dinucleotide ratio (**Figure 3**). The median CpG content for these evolutionarily related viruses is 0.46. However, we noticed that the distribution has a long tail towards even lower values, suggesting that some viruses have a further reduced CpG content (**Figure 3**).

When we plotted the CpG ratio for each virus on a phylogenetic tree (**Figure 4A**), it became clear that the genomes of a subset of bat-associated papillomaviruses have a further decreased CpG content. The order Chiroptera consists of two suborders, Yinpterochiroptera and Yangochiroptera (Lei and Dong 2016; Teeling et al. 2002; Springer et al. 2001). When we associated the viruses with their Yinpterochiroptera and Yangochiroptera hosts, the data demonstrates that the Yangochiropteran papillomavirus genomes have even lower CpG values (orange bars in **Figure 4A**) when compared to the Yinpterochiroptera and other related papillomaviruses in the same phylogenetic clade (blue bars) and members of a closely related clade (grey bars). Indeed, when combined, the Yangochiropteran papillomavirus genomes have a significantly reduced CpG ratio when compared to the other groups (**Figure 4B** ANOVA with posthoc Tukey test).

This reduction in the CpG ratio could be due to an overall lower GC content. We compared the CpG ratio to total genomic GC content (**Figure 4C**). This analysis demonstrates that there is no correlation between the decreased CpG ratio and the total GC content (linear regression: R^2 = 0.008, p-value = 0.27), and the reduced CpG ratio is not simply due to a lower GC content.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The viruses infecting bats in the Yangochiroptera suborder have a reduced CpG content, raising the possibility that the host species is influencing the CpG ratio and evolutionary trajectory of these viruses.

C5. CpG depletion alters codon usage without changing amino acid composition.

The CpG dinucleotide is present in 8 codons coding for five different amino acids. Therefore, reducing CpG dinucleotides is expected to lead to a bias in codon usage or amino acid composition. Roughly 85% of the papillomavirus genome codes for viral proteins (Van Doorslaer et al. 2013; 2017). The viral genome contains several overlapping ORFs (Van Doorslaer 2013; Van Doorslaer and McBride 2016). In some cases, this is a short overlap between the 3' end of one ORF and the 5' end of the downstream ORF (e.g., E6 and E7). In other cases, one ORF is wholly embedded within the coding region of another ORF (e.g., E4 embedded in E2, or E8 within E1). Since these overlapping regions are evolutionarily constrained at multiple codon positions (i.e., codon position 3 in frame 1, would be codon position 2 in the overlapping frame (Miyata and Yasunaga 1978), these overlapping regions were removed from the data (Materials and Methods).

We determined codon usage tables for the non-overlapping coding sequences for each of the papillomaviruses in the subtree described in **Figure 2**. These codon usage tables were compared using the Emboss ‘codcmp' tool to calculate codon usage differences. The more diverse the codon usage, the larger the differences between both tables. This analysis shows that compared to each other, the Yangochiroptera papillomaviruses codon usage is more similar than when other viruses are compared (**Figure 5A**). This suggests that the reduction in CpG leads to a more restricted availability of codons. The relative synonymous codon usage (RSCU) value is the ratio of the observed frequency of one specific synonymous codon to the expected frequency (i.e., no codon usage bias). This ratio is an important measure of codon usage bias (Sharp, Tuohy, and Mosurski 1986). RSCU values higher than 1.6 and lower than 0.6 indicate overrepresented and underrepresented codons, respectively (Wong et al. 2010). CpG containing codons (underlined)

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

are significantly underrepresented in this dataset. For most amino-acids, CpG containing codons are further reduced in the Yangochiroptera papillomaviruses. Of note, in the case of arginine, the CpG containing codons are statistically significantly depleted when compared to the related viruses (**Figure 5C**). Since these codons' CpGs are located in the 1st and 2nd position of the codon, this depletion suggests that non-silent mutations are evolutionarily preferred over maintaining a relatively high CpG content. Of note, despite the biased codon usage, there is no change in the amino acid composition of the different viral proteins (**Figure 5B**). Thus, despite reducing CpG content, the Yangochiroptera papillomaviruses are likely coding for similar proteins. These observations likely explain the more restricted codon usage seen in **Figure 5A**.

C6. Natural selection in the TLR9 of *Yangochiroptera* bats.

Specific residues within the Chiropteran TLR9 have been demonstrated to be under positive selective pressure (Escalera-Zamudio et al. 2015; Jiang et al. 2017). To determine whether diversifying selection differentially affected Yangochiroptera and Yinpterochiroptera TLR9, we constructed a maximum likelihood phylogenetic tree. As was previously reported, bat TLR9 sequences formed a monophyletic clade separate from other eutherian sequences (data not shown), not as a sister-group to carnivores, ungulates, and cetaceans as is seen for other proteins (Tsagkogeorga et al. 2013). We used RELAX (Wertheim et al. 2015), adaptive Branch Site Relative Effect Likelihood (Smith et al. 2015), and Fixed effect Likelihood tests (Kosakovsky Pond and Frost 2005) to detect evidence for evolutionary selection (materials and methods). RELAX demonstrated that evolutionary selection intensified ($K = 6.07$; $LR = 21.05$) within the Yangochiroptera compared to the Yinpterochiroptera. Furthermore, aBSREL recovered evidence for episodic diversifying selection on the branches leading to the Yangochiroptera (**Figure 6B**). Finally, FEL identified 7 sites under diversifying selection within the Yangochiroptera.

We mapped a subset of these residues as well as residues previously identified to be under diversifying selection (Escalera-Zamudio et al. 2015), and sites shown to be functionally important for target recognition (Ohto et al. 2015) onto the structure of TLR9 bound to target DNA (**Figure**

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

6A). Many of the sites are highly variable when compared to the mammalian consensus (**Figure 6B**). Notably, there are apparent differences between the TLR9 sequence of Yangochiroptera and Yinpterochiroptera, specifically within the DNA recognition motif. *In silico* mutation analysis suggests that these Yangochiroptera specific changes would alter how TLR9 recognizes its target DNA. Overall, these mutations lead to a reduction in the positive surface charge of TLR9. Within the Yangochiroptera, K51T leads to the loss of an ionic interaction with the phosphate backbone. The Arg at position 76 is much larger than the canonical His at this position, which will presumably impact the interaction with the DNA. While the P105 residue Vanderwaals bonds with the C6 and T9 residues in the crystalized DNA, the I105 is too bulky to occupy the same conformation as the proline and will likely lead to a loss of the observed bend in the protein at this position. K181 is involved in ionic interactions with the DNA backbone. Q181 has no charge and is too short to interact with the DNA sidechain. E181 will likely charge repel the DNA backbone. Finally, K292 interacts with the DNA backbone, but this interaction in absent in Yangochiroptera TLR9 due to the Ser residue at this position. Overall, the Yangochiroptera TLR9 DNA binding domain is predicted to be functionally different from the Yinpterochiroptera and the other mammalian TLR9 molecules.

C7. CpG depletion points toward a TLR9 signature

Our data demonstrate that the Yangochiroptera TLR9 protein is under selective pressure, and papillomaviruses that infect these hosts have a decreased CpG content. We hypothesized that DNA recognition by TLR9 would lead to a decrease in CpG in the context of a TLR9 specific PAMP. Thus, a specific (set of) tetramers should be depleted within the Yangochiroptera. We calculated the observed/expected ratio for all tetramers and focused on those tetramers with a central CpG (NCGN; **Figure 7A**). This initial analysis indicates that ACGT, GCGT, TCGA, and TCGT are diminished in Yangochiroptera. We calculated the average tetramer ratio for each group of viruses to normalize for the differences in overall CpG content between Yangochiroptera and other viruses, as in **Figure 4**. We compared the proportion of Yang to Yin, Yang to other, and

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Yin to other (**Figure 7B**) for each tetramer. For example, in Yangochiroptera papillomaviruses, 'ACGT' is depleted three- to four-fold compared to other (brown) or Yinpterochiroptera papillomaviruses (orange bar), respectively. Conversely, this tetramer is not depleted when Yinpterochiroptera and other viruses are compared (blue bar). As expected, CpG containing tetramers are depleted in Yangochiroptera papillomaviruses. Using a bootstrap method based on 1000 randomly shuffled sequences (Materials and Methods), the 'ACGT' tetramer was identified as significantly depleted within Yangochiroptera specific viruses (**Figure 7C**). This tetramer is identical to the experimentally validated core mouse TLR9 recognition motif but is different from the human TLR9 PAMP (TCGT) (Pohar et al. 2015; Krieg et al. 1995; Yi et al. 1998; G. Sen et al. 2004; Hartmann and Krieg 2000). This suggests that papillomaviruses associated with Yangochiroptera specifically deplete CpG dinucleotides in the context of a known TLR9 PAMP.

D. Discussion

The data presented here advance our understanding of papillomavirus evolution and host-pathogen interactions. Specifically, we provide evidence that papillomavirus genomes have evolved to avoid detection by TLR9. The implications of this finding for papillomavirus biology are discussed below.

D1. Viruses infecting bats in the suborder Yangochiroptera deplete CpG in a TLR9 dependent manner.

We demonstrate that the genomes of viruses isolated from specific bat species have a highly reduced CpG content. A significant reduction of CpG sites in papillomavirus genomes has been previously documented (Warren, Van Doorslaer, et al. 2015; Upadhyay and Vivekanandan 2015). However, the reason for this depletion is unclear. Of note, in mammalian genomes, CpGs are rare outside of so-called CpG islands (Illingworth and Bird 2009). This is believed to be mainly due to the observation that methylated CpGs are prone to deamination, resulting in C → T mutations, leading to a depletion of CpG sites in the mammalian genomes over evolutionary time.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Our data in **Figure 3** show an increase in TpG and CpA. However, this increase does not appear to be of the same magnitude as the dramatic reduction in CpG seen in the same dataset.

The zinc-finger antiviral protein (ZAP) acts as a broad-spectrum antiviral restriction protein that recognizes CpG rich viral RNA, leading to RNA degradation and inhibition of translation (Gao 2002). Interestingly, it appears that ZAP exploits host CpG suppression to identify non-self RNA. This may explain why multiple RNA viruses have reduced CpG content (Cheng et al. 2013; Greenbaum et al. 2008), independently from CpG methylation as described above (Takata et al. 2017). ZAP was recently shown to restrict the replication of vaccinia virus Ankara (Peng et al. 2020) and HCMV (Lin et al. 2020), demonstrating that ZAP recognizes CpG rich viral RNA and can restrict CpG rich DNA viruses. However, papillomavirus genomes are generally CpG depleted (Warren, Van Doorslaer, et al. 2015; Upadhyay and Vivekanandan 2015). Furthermore, the consensus recognition site for murine ZAP was identified as CN_7GNCG. In this motif, the CG dinucleotide acts as the essential element, while the G further enhances binding affinity 10-fold (Luo et al. 2020). Our tetramer analysis does not identify a downregulation of GNCG in Yangochiroptera specific papillomaviruses (data not shown). Therefore, it appears unlikely that ZAP plays a vital role during papillomavirus infection. However, this would need to be demonstrated experimentally.

In contrast, we provide evidence that the depletion of CpG in papillomavirus genomes is, at least in part, due to the need to avoid detection by TLR9. Unmethylated CpG DNA motifs are recognized by TLR9, leading to an interferon and inflammatory cytokine-mediated antiviral response (Kawai and Akira 2006). By carefully analyzing the CpG content of related Chiropteran papillomaviruses, we demonstrate that viruses isolated from Yangochiroptera have a further decreased CpG content. Importantly, we demonstrate that the Yangochiroptera TLR9 protein is evolving under diversifying selection, specifically sites implicated in DNA recognition. Finally, by analyzing tetramer motifs, we show that Yangochiroptera are specifically depleted in ACGT, a known TLR9 recognition motif. Together these data demonstrate that Yangochiroptera

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

papillomaviruses deplete CpG, in the context of a TLR9 recognition motif, presumably in response to evolutionary changes within the TLR9 protein. This has important implications for papillomavirus biology and evolution.

D2. Recognition of papillomavirus DNA in the endosomes during infectious entry

Shortly after entry, papillomavirus virions are trafficked from early endosomes into acidic late endosome and multivesicular bodies, leading to capsid disassembly and uncoating viral DNA (Campos 2017). Presumably, this exposes the viral DNA to TLR9, leading to an antiviral response. Importantly, TLR9 specifically recognizes unmethylated CpG motifs. Several studies have investigated the methylome of oncogenic human papillomaviruses (Johannsen and Lambert 2013). While these studies have demonstrated that the viral DNA is methylated under specific conditions, it is unknown whether the packaged viral genome contains methylated CpG sites. However, we have some clues that would suggest that viral DNA inside the virion is likely hypomethylated. DNA methyltransferase 1 (DNMT1) is the primary cellular enzyme responsible for maintaining DNA methylation patterns after replication. The DNMT1 protein was found enriched in undifferentiated cells and is reduced as cells differentiate (G. L. Sen et al. 2010, 1). Therefore, it is likely that the reduction in DNMT1 levels will lead to a loss of methylation on the viral genomes destined for packaging and infection of the new tissue.

Differentially methylated CpG dinucleotides are present within consensus E2 binding sites in the viral upstream regulatory region (McBride 2013). The binding of E2 to these binding sites is important for viral replication, transcription, and proper partitioning of the viral genomes to daughter cells (McBride 2013). In many viruses, the full-length E2 protein either activates or represses viral transcription in a dose-dependent manner (Bouvard et al. 1994; Fujii et al. 2001; Thierry and Yaniv 1987; Steger and Corbach 1997). CpG methylation of these sites inhibits E2 binding, presumably altering E2-mediated control of E6/E7 oncogene expression (Thain et al. 1997; Vinokurova and von Knebel Doeberitz 2011). However, the impact of changes to E2BS methylation during cellular differentiation is not understood (Burley, Roberts, and Parish 2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Nonetheless, studies using HPV16 containing cells suggest that the viral URR is hypomethylated upon cellular differentiation (Kim et al. 2003).

A recent study showed that papillomavirus virions package DNA with histones enriched in modifications typically associated with “active” (Porter et al. 2021). Of interest, the authors demonstrate that the levels of H3K4me3 were enriched on virions, compared to cellular controls Conversely, virions were depleted in H3K9me3 (Porter et al. 2021). There is emerging evidence of active associations between histone lysine methylation and DNA methylation (Rose and Klose 2014). For example, MeCP2 binds to methylated CpG (Nan et al. 1998), recruits the Suv39h1/2 histone methyltransferases (Fuks 2003), increasing H3K9me marks (Fuks 2003; Lunyak 2002). In parallel, the H3K9me mark recruits Dnmt3a/b to heterochromatin, leading to *de novo* methylation of CpG sites (Lehnertz et al. 2003; Otani et al. 2009). Since H3K9me3 is depleted in virions, it is tempting to conclude that virion DNA will be hypomethylated. Furthermore, H3K4me3 appears to be mutually exclusive with DNA methylation (M. Weber et al. 2007). H3K4me3 serves as a binding site for H3K9me2 demethylases (Horton et al. 2010), which would lead to loss of DNA methylation. Since virion DNA is enriched for H3K4me3, this further strengthens the hypothesis that viral DNA would be depleted in DNA methylation. Therefore, it seems reasonable to assume that infecting the virus genome will be hypomethylated and therefore serve as a TLR9 PAMP.

As mentioned, recognition by TLR9 would lead to an antiviral response. Indeed, siRNA-mediated knock-down of TLR9 has been shown to dramatically upregulate viral copy number and transcription following infection with HPV16 (Hasan et al. 2013), suggesting that TLR9 can restrict HPV infection. The observation that despite a reduction in viral CpG, HPV16 infection is still improved by interfering with TLR9 (signaling) demonstrates an important rule in host-pathogen interactions. While the loss of all (unmethylated) CpG dinucleotides would avoid detection by TLR9, the virus can likely not completely remove all CpGs from its genome. The virus and the host establish an uneasy balance.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

D3. Sustained flight and the bat immune system

Members of Chiroptera are classified into two suborders – Yinpterochiroptera (Rhinolophoid and megabats) and Yangochiroptera all other bat species (Lei and Dong 2016; Teeling et al. 2002; Springer et al. 2001). These suborders diverged roughly 60 million years ago (Lei and Dong 2016). Interestingly, the Yangochiroptera evolved flight and echolocation simultaneously, while the Yinpterochiroptera evolved these features separately (Anderson and Ruxton 2020). Sustained flight necessitated an increased metabolic capacity (Shen et al. 2010), which required bats to accommodate oxidative metabolism by-products such as DNA damage (Barzilai 2002). Indeed, many genes involved in DNA damage response and immunity have been demonstrated to be under positive evolutionary selection (Hawkins et al. 2019; Zhang et al. 2013). Since both suborders of bats 'invented' flight independently, likely, the corresponding adaptations are also different. Indeed, we and others show that TLR9 is also under diversifying selection, specifically in the Yangochiroptera. While TLR9 likely did not evolve specifically to restrict papillomavirus infections, the virus likely needs to minimize its CpG content.

D4. Direct evidence of co-evolution between the virus and its hosts

Co-evolution alongside their hosts has been suggested to be an essential factor in the evolution of papillomaviruses (Rector et al. 2007; Van Doorslaer 2013). However, the evolutionary history of PVs is complex. PVs isolated from fish form a monophyletic group distinct from those from mammals. However, within the mammalian papillomaviruses, there is no strict codivergence pattern that would unambiguously indicate an ancient relationship between host and virus. There has been no direct evidence in favor of co-evolution between papillomaviruses and their hosts. The co-evolution theory would predict that the virus would need to adapt when the host evolves a new skill. Therefore, as TLR9 evolves new functionalities, the virus would need to respond to preserve the balance between virus and host. Indeed, our data suggest that as Yangochiroptera TLR9 is undergoing diversifying selection, papillomavirus genomes infecting these bats further

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

438 depleted their CpG content, specifically in the context of a known TLR9 PAMP. This is the first
439 direct evidence of co-evolution between this family of viruses and their hosts.
440 In the phylogenetic analysis, the viruses that infect Yinpterochiroptera *and* Yangochiroptera,
441 respectively, are not monophyletic but rather are present in three mixed clades (**Figure 2**; EsPV1,
442 EsPV3, and RfPV1; MscP2 and EhPV1; TbraPV1-3). This suggests that these three main clades
443 diverged before the ancestor of Yinpterochiroptera *and* Yangochiroptera split over 65 million years
444 ago. As these ancestral viruses co-evolved with the Yangochiroptera hosts, they selected for loss
445 of CpG. This occurred at least three separate times in the evolution of Yangochiroptera viruses.
446 This strongly argues against a founder effect but in favor of recurring co-evolutionary interactions.
447 **D5. Immune evasion by nucleotide sequence editing**
448 We previously used computer modeling and reconstruction of ancestral alphapapillomavirus
449 genomes to show that these viruses depleted TpC depletion to allow for replication in tissues with
450 high APOBEC3 expression – presumably to evade restriction by APOBEC3 by selecting for
451 variants that contain reduced target sites in their genomes. We observed a similar correlation
452 between TLR9 and CpG depletion, strengthening the notion that papillomaviruses avoid detection
453 by the immune system by changing the nucleotide composition of their genomes without
454 dramatically changing the protein-coding ability. This strategy likely allows the virus to maintain
455 its core functionalities. Most viral proteins are multifunctional and interact with a plethora of host
456 proteins. Amino acid level changes would likely disrupt these functions. The *de novo* evolution of
457 new proteins is rare and is further complicated by the small genome size and overlapping open
458 reading frames (Van Doorslaer and McBride 2016; Willemsen and Bravo 2019, 5).
459 **D6. Oncogene mediated reduction of TLR9.**
460 We propose that HPVs evade detection by TLR9 in the endosome by depleting CpG dinucleotides
461 from their genomes. Interestingly, the E6 and E7 oncoproteins of different human
462 papillomaviruses have been shown to downregulate the expression of TLR9 (Hasan et al. 2013;
463 2007; Pacini et al. 2015; 2017). Importantly, E6 and E7 are not delivered to the cell during infection

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

but require onset of viral transcription after the viral genome is delivered to the nucleus and presumably has already been sensed in the endosome. This implies that the ability to degrade TLR9 may serve an additional function during the viral lifecycle, independent of initial infection. This idea is supported by the observation that other viruses (Merkel Cell Polyomavirus, Hepatitis B, and EBV) also interfere with TLR9 function during the maintenance phase of the infection (Fathallah et al. 2010; Vincent et al. 2011; Shahzad et al. 2013). Nonetheless, this oncogene-mediated repression of TLR9 occurs after infection and would still necessitate that the virus evades detection during infectious entry.

D7. Conclusion

In conclusion, phylogenetic and genomic analyses of novel bat-associated viruses TbraPV2 and TbraPV3 demonstrate that host-virus interaction, specifically evasion of the innate immune system, affects the evolution of papillomaviruses. These data suggest that TLR9 acts as a restriction factor for papillomavirus infection. Furthermore, we provide the first direct evidence for co-evolution between papillomaviruses and their hosts.

E. Materials and Methods

E1. Data and Code availability

We retrieved full-length reference sequences from the PV database (PaVE; pave.niaid.nih.gov). Data and code for all analyses is available from https://github.com/KVDlab/King-2021. TbraPV2 (MW922427) and TbraPV3 (MW922428) sequences are available on GenBank. Raw sequencing data is available on SRA (PRJNA718335).

E2. Sampling and sample processing

Bats were captured in mist nets set over water sources, extracted from the nets and put in brown paper bags. Bats were held in bags for 20 minutes, removed, and then measured and weighed. Individuals were identified to species in the field using metrics such as forearm length and weight. Feces and urine were collected from the bag or swabbed directly off the bat using a PurFlock 0.14" Ultrafine swab (Puritan, Guilford, Maine). All feces and urine samples were put into tubes

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

containing 0.5 mL buffer consisting of 1x PBS and 50% Glycerol. These samples were held on ice until returning to the lab where they were stored in a -80°C freezer. All applicable international, national and institutional guidelines for the care and use of animals were followed during sampling. The study was approved by the University of Arizona Institutional Animal Care and Use Committee permit #15-583. Permits from the Arizona Department of Game and Fish were numbered SP506475.

Of each of the fecal samples, 5 g was homogenized in SM buffer and the homogenate was centrifuged at 6000 × g for 10 min. The supernatant was sequentially filtered through 0.45 µm and 0.2 µm syringe filters and viral particles in the filtrate were precipitated with 15% (w/v) PEG-8000 with overnight incubation at 4 °C followed by centrifugation at 10,000 ×g as described (Payne et al. 2020). The pellet was resuspended in 500µL of SM Buffer and 200µL of this was used for viral DNA extraction using the High Pure Viral Nucleic Acid Kit (Roche Diagnostics, Indianapolis, IN, USA). The total DNA was amplified using rolling circle amplification (RCA) with the TempliPhi 2000 kit (GE Healthcare, USA) and the RCA products used to prepare Illumina sequencing libraries then sequenced at Novagene Co. Ltd. (Hong Kong) on an Illumina NovaSeq 6000. The paired-end raw reads were trimmed using default settings within Trimmomatic v0.39 (Bolger, Lohse, and Usadel 2014) and the trimmed reads were *de novo* assembled using k-mer values of 33, 66, and 77 within metaSPAdes v 3.12.0 (Bankevich et al. 2012). The resulting contigs greater than 500 nucleotides were analyzed by BLASTx (Altschul et al. 1990) against a local viral protein database constructed from available NCBI RefSeq viral protein sequences (https://ftp.ncbi.nlm.nih.gov/refseq/release/viral/).

E3. Calculation of nucleotide frequencies

To determine a single observed vs. expected (O/E) dinucleotide ratio across the entire viral genome, a custom python script was used that leverages the CompSeq program from Emboss (Warren, Van Doorslaer, et al. 2015). The expected frequencies of dinucleotide 'words' were

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

estimated based on the observed frequency of single bases in the sequences. Only the forward frame was analyzed.

The tetramer content for each genome was calculated as described for the dinucleotides. To normalize tetramer content across groups of viruses we calculated the average O/E ratio across the different groups. These average O/E ratios were compared as indicated in the figure legends. To test whether any of the tested tetramers depletions are statistically significant, we randomly shuffled each viral genome. To ensure that each randomly shuffled sequence would maintain the same dinucleotide ratio as the original sequence, we used the Altschul and Erickson algorithm (Stephen F Altschul and Erickson Bruce W 1985) as implemented by Clote and colleagues (Clote 2005). Based on these shuffled sequences, we calculated the above Yangochiroptera/Yinpterochiroptera ratio. This was repeated 1000 times to establish a null distribution. The 1-percentile was used as a significance cutoff.

E4. Phylogenetic analyses

Annotated sequences (n = 409) were downloaded from the PaVE genome database. A maximum likelihood phylogenetic tree was constructed as described (King and Van Doorslaer 2018). The amino-acid sequences for E1, E2, and L1 of all known papillomaviruses and the new TbraPV2 and TbraPV3 were individually aligned in MAFFT v7.3 (Katoh 2002; 2005) using the L-INS-I algorithm. A partitioning scheme for the concatenated E1-E2-L1 alignment was determined under corrected Akaike information criterion (AICc) implemented in PartitionFInder2 (Lanfear et al. 2017), which separately identified each gene to evolve under the LG+I+G+F evolutionary substitution model. The concatenated E1-E2-L1 alignment was used to infer the best maximum likelihood (ML) phylogenetic tree using RAxML-HPC v.8 (Stamatakis 2014) on CIPRES science gateway (Miller, Pfeiffer, and Schwartz 2010) followed by a rapid bootstrapping analysis. A posteriori bootstopping was automatically rendered in RAxML under the extended majority-rule consensus tree criterion (autoMRE). The best ML tree was rendered and edited in RStudio using the ‘ggtree’ (Yu et al. 2018) and ‘treeio’ (L.-G. Wang et al. 2020) packages.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Taxonomic classification of TbraPV2 and TbraPV3 was based on pairwise sequence identity. The L1 sequence of each pair was aligned at the amino acid level using the L-INS-I algorithm as implemented within the MAFFT v7.3 (Katoh 2002; 2005). This way the alignments preserve the codons. The resulting alignments are back translated to nucleotide alignments and used to calculate pairwise sequence identity.

E5. Coevolution analysis

We used functions in the R 'ape' (Paradis and Schliep 2019) package to extract a well-supported clade from the maximum likelihood phylogenetic tree. The extracted clade represents contains viral sequences in the genera *Lambdapapillomavirus*, *Mupapillomavirus*, *Nupapillomavirus*, *Kappapapillomavirus*, *Sigmapapillomavirus*, and *Dyosigmapapillomavirus*, and the largest set of known of bat papillomaviruses, including the two novel bat papillomaviruses described in this paper. A corresponding host species phylogeny was downloaded from TimeTree (www.timetree.org) (Hedges, Dudley, and Kumar 2006; Kumar et al. 2017; Hedges et al. 2015). A tanglegram representing the evolutionary relationship between the papillomaviruses and their hosts was constructed in the 'phytools' package (Revell 2012). Phytools will optimize the tanglegram by rotating nodes in the rooted phylogenies to minimize crossings between connecting lines between both trees.

An additional subtree was extracted to minimize the impact of the genus *Lamdapapillomavirus*. The viral types included in this smaller dataset are underlined in **Figure 2**. To assess the congruency between PV and host phylogenies, we used the Procrustes Approach to Cophylogenetic Analysis (PACo) (Balbuena, Míguez-Lozano, and Blasco-Costa 2013) as implemented in R for both datasets. Briefly, PACo uses cophenetic distance matrices for the virus and host trees and an association matrix of virus -host interactions. To assess statistical significance, a Procrustean super-imposition of the sum of squared residuals was generated from 1000 network randomizations under the "r2" randomization model. Under this model, host specialization is assumed to drive the virus diversification (Hutchinson et al. 2017). The values

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

for the actual tree comparisons were considered statistically significant if they fell outside the 95% confidence interval (C.I.)

To quantify the similarity between the virus and host phylogenies, we calculated the Wasserstein distance using the 'castor' R package (Louca and Doebeli 2018). The Wasserstein distance is based on a modified graph Laplacian (MGL). The MGL uses evolutionary distances between nodes to construct a matrix which maintains branch length and tree topology information and allows for the comparison of phylogenies from different species. Specifically, the differences between a phylogeny's degree matrix (sum of branch lengths from one node n to all others) and distance matrix (sum of all pairwise branch lengths) is calculated to generate a spectrum of eigenvalues. To calculate a normalized MGL (nMGL) the MGL is divided by the degree matrix. The normalized MGL is specifically useful when comparing trees on different timescales by emphasizing topology over size (Lewitus and Morlon 2016). The Wasserstein distance represents the largest eigenvalue from the spectra of the modified graph Laplacians. All eigenvalues from the graph Laplacian spectrum were used to calculate the Wasserstein tree distance. The Wasserstein tree distance metric calculated in 'castor' considers branch length and tree topology, takes values between 0 and 1. Identical tree topologies would have a Wasserstein distance of 0.

E6. Analysis of codon usage

A custom script was used to delete all overlaps between open reading frames. Briefly, overlaps between E6 and E7, E1 and E8, E2 and E4, L2 and L1 were removed when present. For each overlap, entire codons were removed as not introduce frameshifts. These sequences were concatenated and further analyzed.

Cusp (Emboss suite of tools) was used to generate codon usage tables for each virus. These tables were compared using codcmp (Emboss suite of tools). For each codon in the table codcmp calculates the proportion of a codon to the total number of the codons in the. Next, codcmp calculates the difference between the usage fractions in both tables.

The amino acid composition for each sequence was calculated as described (Carugo 2008).

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

E7. Diversifying selection of analysis of Yangochiroptera TLR9

TLR9 sequences were downloaded from NCBI and translated into putative proteins. The amino acid sequences were aligned using MAFFT v7.3, and back translated into codon-aware nucleotide alignments. FastTree (Price, Dehal, and Arkin 2010) was used to construct a maximum likelihood phylogenetic tree using the GTR substitution model of evolution.

To determine whether the strength of natural selection intensified along the along Yangochiroptera compared to the Yinpterochiroptera, we used RELAX. After fitting a codon model with three ω classes to the phylogeny (null model), RELAX then tests for changes to the intensity of selection by introducing a selection parameter k. The null and alternative models are compared using a Likelihood Ratio Test. A significant result of k>1 indicates that selection strength has been intensified along the test branches (Wertheim et al. 2015).

aBSREL (adaptive Branch-Site Random Effects Likelihood) was used to test if positive selection has occurred on the branches leading to Yangochiroptera. aBSREL determines whether a proportion of sites have evolved under positive selection (Smith et al. 2015).

Finally, FEL (Fixed Effects Likelihood) was used to infer non-synonymous (dN) and synonymous (dS) substitution rates on a per-site basis. This method assumes that the selection pressure for each site is constant along the entire phylogeny. FEL fits a MG94xREV model to each codon site to infer nonsynonymous and synonymous substitution rates at each site. A Likelihood Ratio Test determines if dN is significantly greater than dS.

E8. Statistical analysis

One- or two-way analysis of variance (ANOVA) were used where appropriate. Data are presented as box-and-whisker plots with Tukey's method for outliers noted as distinct data points. All graphs were generated using R. Results were considered statistically significant at a *P*-value of < 0.05.

References

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

618 Akira, Shizuo, Satoshi Uematsu, and Osamu Takeuchi. 2006. "Pathogen Recognition and Innate
619 Immunity." *Cell* 124 (4): 783–801. https://doi.org/10.1016/j.cell.2006.02.015.
620 Aksoy, Pinar, Elinor Y. Gottschalk, and Patricio I. Meneses. 2017. "HPV Entry into Cells."
621 *Mutation Research. Reviews in Mutation Research* 772 (June): 13–22.
622 https://doi.org/10.1016/j.mrrev.2016.09.004.
623 Alexopoulou, Lena, Agnieszka Czopik Holt, Ruslan Medzhitov, and Richard A. Flavell. 2001.
624 "Recognition of Double-Stranded RNA and Activation of NF-KB by Toll-like Receptor
625 3." *Nature* 413 (6857): 732–38. https://doi.org/10.1038/35099560.
626 Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. "Basic Local
627 Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10.
628 https://doi.org/10.1016/S0022-2836(05)80360-2.
629 Anderson, Sophia C., and Graeme D. Ruxton. 2020. "The Evolution of Flight in Bats: A Novel
630 Hypothesis." *Mammal Review* 50 (4): 426–39. https://doi.org/10.1111/mam.12211.
631 Aydin, Inci, Ruth Villalonga-Planells, Lilo Greune, Matthew P. Bronnimann, Christine M.
632 Calton, Miriam Becker, Kun-Yi Lai, Samuel K. Campos, M. Alexander Schmidt, and
633 Mario Schelhaas. 2017. "A Central Region in the Minor Capsid Protein of
634 Papillomaviruses Facilitates Viral Genome Tethering and Membrane Penetration for
635 Mitotic Nuclear Entry." *PLoS Pathogens* 13 (5): e1006308.
636 https://doi.org/10.1371/journal.ppat.1006308.
637 Aydin, Inci, Susanne Weber, Berend Snijder, Pilar Samperio Ventayol, Andreas Kühbacher,
638 Miriam Becker, Patricia M. Day, et al. 2014. "Large Scale RNAi Reveals the
639 Requirement of Nuclear Envelope Breakdown for Nuclear Import of Human
640 Papillomaviruses." *PLoS Pathogens* 10 (5): e1004162.
641 https://doi.org/10.1371/journal.ppat.1004162.
642 Balbuena, Juan Antonio, Raúl Míguez-Lozano, and Isabel Blasco-Costa. 2013. "PACo: A Novel
643 Procrustes Application to Cophylogenetic Analysis." *PloS One* 8 (4): e61048.
644 https://doi.org/10.1371/journal.pone.0061048.
645 Banerjee, Arinjay, Michelle L. Baker, Kirsten Kulcsar, Vikram Misra, Raina Plowright, and
646 Karen Mossman. 2020. "Novel Insights Into Immune Systems of Bats." *Frontiers in*
647 *Immunology* 11: 26. https://doi.org/10.3389/fimmu.2020.00026.
648 Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin,
649 Alexander S. Kulikov, Valery M. Lesin, et al. 2012. "SPAdes: A New Genome Assembly
650 Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational*
651 *Biology* 19 (5): 455–77. https://doi.org/10.1089/cmb.2012.0021.
652 Barzilai, A. 2002. "ATM Deficiency and Oxidative Stress: A New Dimension of Defective
653 Response to DNA Damage." *DNA Repair* 1 (1): 3–25. https://doi.org/10.1016/S1568-
654 7864(01)00007-6.
655 Bell, J. K., J. Askins, P. R. Hall, D. R. Davies, and D. M. Segal. 2006. "The DsRNA Binding
656 Site of Human Toll-like Receptor 3." *Proceedings of the National Academy of Sciences*
657 103 (23): 8792–97. https://doi.org/10.1073/pnas.0603245103.
658 Bernard, Hans-Ulrich, Robert D. Burk, Zigui Chen, Koenraad van Doorslaer, Harald zur Hausen,
659 and Ethel-Michele de Villiers. 2010. "Classification of Papillomaviruses (PVs) Based on
660 189 PV Types and Proposal of Taxonomic Amendments." *Virology* 401 (1): 70–79.
661 https://doi.org/10.1016/j.virol.2010.02.002.
662 Bieback, Karen, Egil Lien, Ingo M. Klagge, Elita Avota, Jürgen Schneider-Schaulies, W. Paul
663 Duprex, Herrmann Wagner, Carsten J. Kirschning, Volker ter Meulen, and Sibylle

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

664 Schneider-Schaulies. 2002. “Hemagglutinin Protein of Wild-Type Measles Virus
665 Activates Toll-Like Receptor 2 Signaling.” *Journal of Virology* 76 (17): 8729–36.
666 https://doi.org/10.1128/JVI.76.17.8729-8736.2002.
667 Blanco, Jorge C. G., Marina S. Boukhvalova, Kari A. Shirey, Gregory A. Prince, and Stefanie N.
668 Vogel. 2010. “New Insights for Development of a Safe and Protective RSV Vaccine.”
669 *Human Vaccines* 6 (6): 482–92. https://doi.org/10.4161/hv.6.6.11562.
670 Boehme, Karl W., Mario Guerrero, and Teresa Compton. 2006. “Human Cytomegalovirus
671 Envelope Glycoproteins B and H Are Necessary for TLR2 Activation in Permissive
672 Cells.” *The Journal of Immunology* 177 (10): 7094–7102.
673 https://doi.org/10.4049/jimmunol.177.10.7094.
674 Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A Flexible Trimmer
675 for Illumina Sequence Data.” *Bioinformatics* 30 (15): 2114–20.
676 https://doi.org/10.1093/bioinformatics/btu170.
677 Bouvard, V., A. Storey, D. Pim, and L. Banks. 1994. “Characterization of the Human
678 Papillomavirus E2 Protein: Evidence of Trans-Activation and Trans-Repression in
679 Cervical Keratinocytes.” *The EMBO Journal* 13 (22): 5451–59.
680 Bowie, Andrew G., and Leonie Unterholzner. 2008. “Viral Evasion and Subversion of Pattern-
681 Recognition Receptor Signalling.” *Nature Reviews. Immunology* 8 (12): 911–22.
682 https://doi.org/10.1038/nri2436.
683 Brook, Cara E., and Andrew P. Dobson. 2015. “Bats as ‘Special’ Reservoirs for Emerging
684 Zoonotic Pathogens.” *Trends in Microbiology* 23 (3): 172–80.
685 https://doi.org/10.1016/j.tim.2014.12.004.
686 Buck, Christopher B., Patricia M. Day, and Benes L. Trus. 2013. “The Papillomavirus Major
687 Capsid Protein L1.” *Virology* 445 (1–2): 169–74.
688 https://doi.org/10.1016/j.virol.2013.05.038.
689 Buck, Christopher B., Koenraad Van Doorslaer, Alberto Peretti, Eileen M. Geoghegan, Michael
690 J. Tisza, Ping An, Joshua P. Katz, et al. 2016. “The Ancient Evolutionary History of
691 Polyomaviruses.” *PLoS Pathogens* 12 (4): e1005574.
692 https://doi.org/10.1371/journal.ppat.1005574.
693 Burley, Megan, Sally Roberts, and Joanna L. Parish. 2020. “Epigenetic Regulation of Human
694 Papillomavirus Transcription in the Productive Virus Life Cycle.” *Seminars in*
695 *Immunopathology* 42 (2): 159–71. https://doi.org/10.1007/s00281-019-00773-0.
696 Bushnell, Brian. 2014. “BBMap: A Fast, Accurate, Splice-Aware Aligner.”
697 https://www.osti.gov/biblio/1241166.
698 Calton, Christine M., Matthew P. Bronnimann, Ariana R. Manson, Shuaizhi Li, Janice A.
699 Chapman, Marcela Suarez-Berumen, Tatum R. Williamson, Sudheer K. Molugu, Ricardo
700 A. Bernal, and Samuel K. Campos. 2017. “Translocation of the Papillomavirus
701 L2/VDNA Complex across the Limiting Membrane Requires the Onset of Mitosis.”
702 *PLoS Pathogens* 13 (5): e1006200. https://doi.org/10.1371/journal.ppat.1006200.
703 Campos, Samuel K. 2017. “Subcellular Trafficking of the Papillomavirus Genome during Initial
704 Infection: The Remarkable Abilities of Minor Capsid Protein L2.” *Viruses* 9 (12).
705 https://doi.org/10.3390/v9120370.
706 Carey, Clayton M., Apurva A. Govande, Juliane M. Cooper, Melissa K. Hartley, Philip J.
707 Kranzusch, and Nels C. Elde. 2019. “Recurrent Loss-of-Function Mutations Reveal Costs
708 to OAS1 Antiviral Activity in Primates.” *Cell Host & Microbe* 25 (2): 336-343.e4.
709 https://doi.org/10.1016/j.chom.2019.01.001.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

710 Carugo, Oliviero. 2008. “Amino Acid Composition and Protein Dimension.” *Protein Science: A*
711 *Publication of the Protein Society* 17 (12): 2187–91.
712 https://doi.org/10.1110/ps.037762.108.
713 Cerqueira, Carla, Pilar Samperio Ventayol, Christian Vogeley, and Mario Schelhaas. 2015.
714 “Kallikrein-8 Proteolytically Processes Human Papillomaviruses in the Extracellular
715 Space To Facilitate Entry into Host Cells.” *Journal of Virology* 89 (14): 7038–52.
716 https://doi.org/10.1128/JVI.00234-15.
717 Cheng, Xiaofei, Nasar Virk, Wei Chen, Shuqin Ji, Shuxian Ji, Yuqiang Sun, and Xiaoyun Wu.
718 2013. “CpG Usage in RNA Viruses: Data and Hypotheses.” Edited by Robert D. Burk.
719 *PLoS ONE* 8 (9): e74109. https://doi.org/10.1371/journal.pone.0074109.
720 Choe, J. 2005. “Crystal Structure of Human Toll-Like Receptor 3 (TLR3) Ectodomain.” *Science*
721 309 (5734): 581–85. https://doi.org/10.1126/science.1115253.
722 Clote, P. 2005. “Structural RNA Has Lower Folding Energy than Random RNA of the Same
723 Dinucleotide Frequency.” *RNA* 11 (5): 578–91. https://doi.org/10.1261/rna.7220505.
724 Day, Patricia M., Carl C. Baker, Douglas R. Lowy, and John T. Schiller. 2004. “Establishment of
725 Papillomavirus Infection Is Enhanced by Promyelocytic Leukemia Protein (PML)
726 Expression.” *Proceedings of the National Academy of Sciences of the United States of*
727 *America* 101 (39): 14252–57. https://doi.org/10.1073/pnas.0404229101.
728 Day, Patricia M., and Mario Schelhaas. 2014. “Concepts of Papillomavirus Entry into Host
729 Cells.” *Current Opinion in Virology* 4 (February): 24–31.
730 https://doi.org/10.1016/j.coviro.2013.11.002.
731 Day, Patricia M., Cynthia D. Thompson, Rachel M. Schowalter, Douglas R. Lowy, and John T.
732 Schiller. 2013. “Identification of a Role for the Trans-Golgi Network in Human
733 Papillomavirus 16 Pseudovirus Infection.” *Journal of Virology* 87 (7): 3862–70.
734 https://doi.org/10.1128/JVI.03222-12.
735 Diebold, S. S. 2004. “Innate Antiviral Responses by Means of TLR7-Mediated Recognition of
736 Single-Stranded RNA.” *Science* 303 (5663): 1529–31.
737 https://doi.org/10.1126/science.1093616.
738 DiGiuseppe, Stephen, Malgorzata Bienkowska-Haba, Lucile G. M. Guion, Timothy R. Keiffer,
739 and Martin Sapp. 2017. “Human Papillomavirus Major Capsid Protein L1 Remains
740 Associated with the Incoming Viral Genome throughout the Entry Process.” *Journal of*
741 *Virology* 91 (16). https://doi.org/10.1128/JVI.00537-17.
742 DiGiuseppe, Stephen, Malgorzata Bienkowska-Haba, Lucile G. Guion, and Martin Sapp. 2017.
743 “Cruising the Cellular Highways: How Human Papillomavirus Travels from the Surface
744 to the Nucleus.” *Virus Research* 231 (March): 1–9.
745 https://doi.org/10.1016/j.virusres.2016.10.015.
746 Escalera-Zamudio, Marina, M. Lisandra Zepeda-Mendoza, Elizabeth Loza-Rubio, Edith Rojas-
747 Anaya, Maria L. Méndez-Ojeda, Carlos F. Arias, and Alex D. Greenwood. 2015. “The
748 Evolution of Bat Nucleic Acid-Sensing Toll-like Receptors.” *Molecular Ecology* 24 (23):
749 5899–5909. https://doi.org/10.1111/mec.13431.
750 Fathallah, Ikbal, Peggy Parroche, Henri Gruffat, Claudia Zannetti, Hanna Johansson, Jiping Yue,
751 Evelyn Manet, Massimo Tommasino, Bakary S. Sylla, and Uzma A. Hasan. 2010. “EBV
752 Latent Membrane Protein 1 Is a Negative Regulator of TLR9.” *Journal of Immunology*
753 *(Baltimore, Md.: 1950)* 185 (11): 6439–47. https://doi.org/10.4049/jimmunol.0903459.
754 Fujii, Takuma, Janet L. Brandsma, Xueyan Peng, Srinivasan Srimatkandada, Lei Li, Allon
755 Canaan, and Albert B. Deisseroth. 2001. “High and Low Levels of Cottontail Rabbit

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Papillomavirus E2 Protein Generate Opposite Effects on Gene Expression." *Journal of Biological Chemistry* 276 (2): 867–74. https://doi.org/10.1074/jbc.M007120200.

Fuks, F. 2003. "The DNA Methyltransferases Associate with HP1 and the SUV39H1 Histone Methyltransferase." *Nucleic Acids Research* 31 (9): 2305–12. https://doi.org/10.1093/nar/gkg332.

Gao, G. 2002. "Inhibition of Retroviral RNA Production by ZAP, a CCCH-Type Zinc Finger Protein." *Science* 297 (5587): 1703–6. https://doi.org/10.1126/science.1074276.

Gorbunova, Vera, Andrei Seluanov, and Brian K. Kennedy. 2020. "The World Goes Bats: Living Longer and Tolerating Viruses." *Cell Metabolism* 32 (1): 31–43. https://doi.org/10.1016/j.cmet.2020.06.013.

Gottschling, Marc, Markus Göker, Alexandros Stamatakis, Olaf R. P. Bininda-Emonds, Ingo Nindl, and Ignacio G. Bravo. 2011. "Quantifying the Phylodynamic Forces Driving Papillomavirus Evolution." *Molecular Biology and Evolution* 28 (7): 2101–13. https://doi.org/10.1093/molbev/msr030.

Gottschling, Marc, Alexandros Stamatakis, Ingo Nindl, Eggert Stockfleth, Angel Alonso, and Ignacio G. Bravo. 2007. "Multiple Evolutionary Mechanisms Drive Papillomavirus Diversification." *Molecular Biology and Evolution* 24 (5): 1242–58. https://doi.org/10.1093/molbev/msm039.

Gowen, Brian B., Justin D. Hoopes, Min-Hui Wong, Kie-Hoon Jung, Kevin C. Isakson, Lena Alexopoulou, Richard A. Flavell, and Robert W. Sidwell. 2006. "TLR3 Deletion Limits Mortality and Disease Severity Due to Phlebovirus Infection." *The Journal of Immunology* 177 (9): 6301–7. https://doi.org/10.4049/jimmunol.177.9.6301.

Greenbaum, Benjamin D., Arnold J. Levine, Gyan Bhanot, and Raul Rabadan. 2008. "Patterns of Evolution and Host Gene Mimicry in Influenza and Other RNA Viruses." Edited by Edward C. Holmes. *PLoS Pathogens* 4 (6): e1000079. https://doi.org/10.1371/journal.ppat.1000079.

Gupta, Chhedi Lal, Salman Akhtar, Andrew Waye, Nihar R. Pandey, Neelam Pathak, and Preeti Bajpai. 2015. "Cross Talk between Leishmania Donovani CpG DNA and Toll-like Receptor 9: An Immunoinformatics Approach." *Biochemical and Biophysical Research Communications* 459 (3): 424–29. https://doi.org/10.1016/j.bbrc.2015.02.121.

Hartmann, Gunther, and Arthur M. Krieg. 2000. "Mechanism and Function of a Newly Identified CpG DNA Motif in Human Primary B Cells." *The Journal of Immunology* 164 (2): 944–53. https://doi.org/10.4049/jimmunol.164.2.944.

Hasan, Uzma A., Elizabeth Bates, Fumihiko Takeshita, Alexandra Biliato, Rosita Accardi, Veronique Bouvard, Mariam Mansour, et al. 2007. "TLR9 Expression and Function Is Abolished by the Cervical Cancer-Associated Human Papillomavirus Type 16." *The Journal of Immunology* 178 (5): 3186–97. https://doi.org/10.4049/jimmunol.178.5.3186.

Hasan, Uzma A., Claudia Zannetti, Peggy Parroche, Nadège Goutagny, Marine Malfroy, Guillaume Roblot, Christine Carreira, et al. 2013. "The Human Papillomavirus Type 16 E7 Oncoprotein Induces a Transcriptional Repressor Complex on the Toll-like Receptor 9 Promoter." *Journal of Experimental Medicine* 210 (7): 1369–87. https://doi.org/10.1084/jem.20122394.

Hawkins, John A., Maria E. Kaczmarek, Marcel A. Müller, Christian Drosten, William H. Press, and Sara L. Sawyer. 2019. "A Metaanalysis of Bat Phylogenetics and Positive Selection Based on Genomes and Transcriptomes from 18 Species." *Proceedings of the National Academy of Sciences* 116 (23): 11351–60. https://doi.org/10.1073/pnas.1814995116.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

802 Hedges, S. Blair, Joel Dudley, and Sudhir Kumar. 2006. "TimeTree: A Public Knowledge-Base
803 of Divergence Times among Organisms." *Bioinformatics (Oxford, England)* 22 (23):
804 2971–72. https://doi.org/10.1093/bioinformatics/btl505.
805 Hedges, S. Blair, Julie Marin, Michael Suleski, Madeline Paymer, and Sudhir Kumar. 2015.
806 "Tree of Life Reveals Clock-like Speciation and Diversification." *Molecular Biology and*
807 *Evolution* 32 (4): 835–45. https://doi.org/10.1093/molbev/msv037.
808 Heil, F. 2004. "Species-Specific Recognition of Single-Stranded RNA via Toll-like Receptor 7
809 and 8." *Science* 303 (5663): 1526–29. https://doi.org/10.1126/science.1093620.
810 Hemmi, Hiroaki, Tsuneyasu Kaisho, Osamu Takeuchi, Shintaro Sato, Hideki Sanjo, Katsuaki
811 Hoshino, Takao Horiuchi, Hideyuki Tomizawa, Kiyoshi Takeda, and Shizuo Akira. 2002.
812 "Small Anti-Viral Compounds Activate Immune Cells via the TLR7 MyD88-Dependent
813 Signaling Pathway." *Nature Immunology* 3 (2): 196–200. https://doi.org/10.1038/ni758.
814 Horton, John R, Anup K Upadhyay, Hank H Qi, Xing Zhang, Yang Shi, and Xiaodong Cheng.
815 2010. "Enzymatic and Structural Insights for Substrate Specificity of a Family of Jumonji
816 Histone Lysine Demethylases." *Nature Structural & Molecular Biology* 17 (1): 38–43.
817 https://doi.org/10.1038/nsmb.1753.
818 Hutchinson, Matthew C., E. Fernando Cagua, Juan A. Balbuena, Daniel B. Stouffer, and
819 Timothée Poisot. 2017. "Paco: Implementing Procrustean Approach to Cophylogeny in
820 R." Edited by Richard Fitzjohn. *Methods in Ecology and Evolution* 8 (8): 932–40.
821 https://doi.org/10.1111/2041-210X.12736.
822 Illingworth, Robert S., and Adrian P. Bird. 2009. "CpG Islands - 'A Rough Guide.'" *FEBS*
823 *Letters* 583 (11): 1713–20. https://doi.org/10.1016/j.febslet.2009.04.012.
824 Jiang, Haiying, Juan Li, Linmiao Li, Xiujuan Zhang, Lihong Yuan, and Jinping Chen. 2017.
825 "Selective Evolution of Toll-like Receptors 3, 7, 8, and 9 in Bats." *Immunogenetics* 69
826 (4): 271–85. https://doi.org/10.1007/s00251-016-0966-2.
827 Johannsen, Eric, and Paul F. Lambert. 2013. "Epigenetics of Human Papillomaviruses." *Virology*
828 445 (1–2): 205–12. https://doi.org/10.1016/j.virol.2013.07.016.
829 Jude, Brooke A, Yelena Pobezinskaya, Jennifer Bishop, Susannah Parke, Ruslan M Medzhitov,
830 Alexander V Chervonsky, and Tatyana V Golovkina. 2003. "Subversion of the Innate
831 Immune System by a Retrovirus." *Nature Immunology* 4 (6): 573–78.
832 https://doi.org/10.1038/ni926.
833 Jurk, Marion, Florian Heil, Jörg Vollmer, Christian Schetter, Arthur M. Krieg, Hermann Wagner,
834 Grayson Lipford, and Stefan Bauer. 2002. "Human TLR7 or TLR8 Independently Confer
835 Responsiveness to the Antiviral Compound R-848." *Nature Immunology* 3 (6): 499.
836 https://doi.org/10.1038/ni0602-499.
837 Katoh, K. 2002. "MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on
838 Fast Fourier Transform." *Nucleic Acids Research* 30 (14): 3059–66.
839 https://doi.org/10.1093/nar/gkf436.
840 ———. 2005. "MAFFT Version 5: Improvement in Accuracy of Multiple Sequence
841 Alignment." *Nucleic Acids Research* 33 (2): 511–18. https://doi.org/10.1093/nar/gki198.
842 Kawai, Taro, and Shizuo Akira. 2006. "Innate Immune Recognition of Viral Infection." *Nature*
843 *Immunology* 7 (2): 131–37. https://doi.org/10.1038/ni1303.
844 Kim, Kitai, Peggy A. Garner-Hamrick, Chris Fisher, Denis Lee, and Paul F. Lambert. 2003.
845 "Methylation Patterns of Papillomavirus DNA, Its Influence on E2 Function, and
846 Implications in Viral Infection." *Journal of Virology* 77 (23): 12450–59.
847 https://doi.org/10.1128/jvi.77.23.12450-12459.2003.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

King, Kelly M., and Koenraad Van Doorslaer. 2018. "Building (Viral) Phylogenetic Trees Using a Maximum Likelihood Approach." *Current Protocols in Microbiology* 51 (1): e63. https://doi.org/10.1002/cpmc.63.

Kosakovsky Pond, Sergei L., and Simon D. W. Frost. 2005. "Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection." *Molecular Biology and Evolution* 22 (5): 1208–22. https://doi.org/10.1093/molbev/msi105.

Krieg, Arthur M., Ae-Kyung Yi, Sara Matson, Thomas J. Waldschmidt, Gail A. Bishop, Rebecca Teasdale, Gary A. Koretzky, and Dennis M. Klinman. 1995. "CpG Motifs in Bacterial DNA Trigger Direct B-Cell Activation." *Nature* 374 (6522): 546–49. https://doi.org/10.1038/374546a0.

Kumar, Sudhir, Glen Stecher, Michael Suleski, and S. Blair Hedges. 2017. "TimeTree: A Resource for Timelines, Timetrees, and Divergence Times." *Molecular Biology and Evolution* 34 (7): 1812–19. https://doi.org/10.1093/molbev/msx116.

Lanfear, Robert, Paul B. Frandsen, April M. Wright, Tereza Senfeld, and Brett Calcott. 2017. "PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses." *Molecular Biology and Evolution* 34 (3): 772–73. https://doi.org/10.1093/molbev/msw260.

Lehnertz, Bernhard, Yoshihide Ueda, Alwin A.H.A. Derijck, Ulrich Braunschweig, Laura Perez-Burgos, Stefan Kubicek, Taiping Chen, En Li, Thomas Jenuwein, and Antoine H.F.M. Peters. 2003. "Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin." *Current Biology* 13 (14): 1192–1200. https://doi.org/10.1016/S0960-9822(03)00432-9.

Lei, Ming, and Dong Dong. 2016. "Phylogenomic Analyses of Bat Subordinal Relationships Based on Transcriptome Data." *Scientific Reports* 6 (1): 27726. https://doi.org/10.1038/srep27726.

Lewitus, Eric, and Helene Morlon. 2016. "Characterizing and Comparing Phylogenies from Their Laplacian Spectrum." *Systematic Biology* 65 (3): 495–507. https://doi.org/10.1093/sysbio/syv116.

Lin, Yao-Tang, Stephen Chiweshe, Dominique McCormick, Anna Raper, Arthur Wickenhagen, Victor DeFillipis, Eleanor Gaunt, Peter Simmonds, Sam J. Wilson, and Finn Grey. 2020. "Human Cytomegalovirus Evades ZAP Detection by Suppressing CpG Dinucleotides in the Major Immediate Early 1 Gene." Edited by Eain A. Murphy. *PLOS Pathogens* 16 (9): e1008844. https://doi.org/10.1371/journal.ppat.1008844.

Lipovsky, Alex, Andreea Popa, Genaro Pimienta, Michael Wyler, Ashima Bhan, Leena Kuruvilla, Marie-Aude Guie, et al. 2013. "Genome-Wide SiRNA Screen Identifies the Retromer as a Cellular Entry Factor for Human Papillomavirus." *Proceedings of the National Academy of Sciences of the United States of America* 110 (18): 7452–57. https://doi.org/10.1073/pnas.1302164110.

Louca, Stilianos, and Michael Doebeli. 2018. "Efficient Comparative Phylogenetics on Large Trees." *Bioinformatics (Oxford, England)* 34 (6): 1053–55. https://doi.org/10.1093/bioinformatics/btx701.

Lunyak, V. V. 2002. "Corepressor-Dependent Silencing of Chromosomal Regions Encoding Neuronal Genes." *Science* 298 (5599): 1747–52. https://doi.org/10.1126/science.1076469.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Luo, Xiu, Xinlu Wang, Yina Gao, Jingpeng Zhu, Songqing Liu, Guangxia Gao, and Pu Gao. 2020. "Molecular Mechanism of RNA Recognition by Zinc-Finger Antiviral Protein." *Cell Reports* 30 (1): 46-52.e4. https://doi.org/10.1016/j.celrep.2019.11.116.

McBride, Alison A. 2013. "The Papillomavirus E2 Proteins." *Virology* 445 (1–2): 57–79. https://doi.org/10.1016/j.virol.2013.06.006.

Medzhitov, Ruslan. 2007. "Recognition of Microorganisms and Activation of the Immune Response." *Nature* 449 (7164): 819–26. https://doi.org/10.1038/nature06246.

Miller, Mark A., Wayne Pfeiffer, and Terri Schwartz. 2010. "Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees." In *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. New Orleans, LA, USA: IEEE. https://doi.org/10.1109/GCE.2010.5676129.

Miyata, T., and T. Yasunaga. 1978. "Evolution of Overlapping Genes." *Nature* 272 (5653): 532–35. https://doi.org/10.1038/272532a0.

Murawski, Matthew R., Glennice N. Bowen, Anna M. Cerny, Larry J. Anderson, Lia M. Haynes, Ralph A. Tripp, Evelyn A. Kurt-Jones, and Robert W. Finberg. 2009. "Respiratory Syncytial Virus Activates Innate Immunity through Toll-Like Receptor 2." *Journal of Virology* 83 (3): 1492–1500. https://doi.org/10.1128/JVI.00671-08.

Nan, X., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird. 1998. "Transcriptional Repression by the Methyl-CpG-Binding Protein MeCP2 Involves a Histone Deacetylase Complex." *Nature* 393 (6683): 386–89. https://doi.org/10.1038/30764.

Ohto, Umeharu, Takuma Shibata, Hiromi Tanji, Hanako Ishida, Elena Krayukhina, Susumu Uchiyama, Kensuke Miyake, and Toshiyuki Shimizu. 2015. "Structural Basis of CpG and Inhibitory DNA Recognition by Toll-like Receptor 9." *Nature* 520 (7549): 702–5. https://doi.org/10.1038/nature14138.

O'Shea, Thomas J., Paul M. Cryan, Andrew A. Cunningham, Anthony R. Fooks, David T.S. Hayman, Angela D. Luis, Alison J. Peel, Raina K. Plowright, and James L.N. Wood. 2014. "Bat Flight and Zoonotic Viruses." *Emerging Infectious Diseases* 20 (5): 741–45. https://doi.org/10.3201/eid2005.130539.

Oshiumi, Hiroyuki, Masaaki Okamoto, Ken Fujii, Takashi Kawanishi, Misako Matsumoto, Satoshi Koike, and Tsukasa Seya. 2011. "The TLR3/TICAM-1 Pathway Is Mandatory for Innate Immune Responses to Poliovirus Infection." *The Journal of Immunology* 187 (10): 5320–27. https://doi.org/10.4049/jimmunol.1101503.

Otani, Junji, Toshiyuki Nankumo, Kyohei Arita, Susumu Inamoto, Mariko Ariyoshi, and Masahiro Shirakawa. 2009. "Structural Basis for Recognition of H3K4 Methylation Status by the DNA Methyltransferase 3A ATRX-DNMT3-DNMT3L Domain." *EMBO Reports* 10 (11): 1235–41. https://doi.org/10.1038/embor.2009.218.

Pacini, Laura, Maria Grazia Ceraolo, Assunta Venuti, Giusi Melita, Uzma A. Hasan, Rosita Accardi, and Massimo Tommasino. 2017. "UV Radiation Activates Toll-Like Receptor 9 Expression in Primary Human Keratinocytes, an Event Inhibited by Human Papillomavirus 38 E6 and E7 Oncoproteins." *Journal of Virology* 91 (19). https://doi.org/10.1128/JVI.01123-17.

Pacini, Laura, Claudia Savini, Raffaella Ghittoni, Djamel Saidj, Jerome Lamartine, Uzma A. Hasan, Rosita Accardi, and Massimo Tommasino. 2015. "Downregulation of Toll-Like Receptor 9 Expression by Beta Human Papillomavirus 38 and Implications for Cell Cycle

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Control." Edited by R. M. Sandri-Goldin. *Journal of Virology* 89 (22): 11396–405. https://doi.org/10.1128/JVI.02151-15.

Paradis, Emmanuel, and Klaus Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics (Oxford, England)* 35 (3): 526–28. https://doi.org/10.1093/bioinformatics/bty633.

Payne, Natalie, Simona Kraberger, Rafaela S. Fontenele, Kara Schmidlin, Melissa H. Bergeman, Ivonne Cassaigne, Melanie Culver, Arvind Varsani, and Koenraad Van Doorslaer. 2020. "Novel Circoviruses Detected in Feces of Sonoran Felids." *Viruses* 12 (9). https://doi.org/10.3390/v12091027.

Peng, Chen, Linda S. Wyatt, Shira G. Glushakow-Smith, Madhu Lal-Nag, Andrea S. Weisberg, and Bernard Moss. 2020. "Zinc-Finger Antiviral Protein (ZAP) Is a Restriction Factor for Replication of Modified Vaccinia Virus Ankara (MVA) in Human Cells." Edited by Matthew S. Wiebe. *PLOS Pathogens* 16 (8): e1008845. https://doi.org/10.1371/journal.ppat.1008845.

Pohar, Jelka, Duško Lainšček, Ryutaro Fukui, Chikako Yamamoto, Kensuke Miyake, Roman Jerala, and Mojca Benčina. 2015. "Species-Specific Minimal Sequence Motif for Oligodeoxyribonucleotides Activating Mouse TLR9." *The Journal of Immunology* 195 (9): 4396–4405. https://doi.org/10.4049/jimmunol.1500600.

Popa, Andreea, Wei Zhang, Megan S. Harrison, Kylia Goodner, Teymur Kazakov, Edward C. Goodwin, Alex Lipovsky, Christopher G. Burd, and Daniel DiMaio. 2015. "Direct Binding of Retromer to Human Papillomavirus Type 16 Minor Capsid Protein L2 Mediates Endosome Exit during Viral Infection." *PLoS Pathogens* 11 (2): e1004699. https://doi.org/10.1371/journal.ppat.1004699.

Porter, Samuel S., Jennifer C. Liddle, Kristen Browne, Diana V. Pastrana, Benjamin A. Garcia, Christopher B. Buck, Matthew D. Weitzman, and Alison A. McBride. 2021. "Histone Modifications in Papillomavirus Virion Minichromosomes." Edited by Thomas Shenk. *MBio* 12 (1): e03274-20, /mbio/12/1/mBio.03274-20.atom. https://doi.org/10.1128/mBio.03274-20.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. "FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments." Edited by Art F. Y. Poon. *PLoS ONE* 5 (3): e9490. https://doi.org/10.1371/journal.pone.0009490.

Rassa, John C., Jennifer L. Meyers, Yuanming Zhang, Rama Kudaravalli, and Susan R. Ross. 2002. "Murine Retroviruses Activate B Cells via Interaction with Toll-like Receptor 4." *Proceedings of the National Academy of Sciences of the United States of America* 99 (4): 2281–86. https://doi.org/10.1073/pnas.042355399.

Rector, Annabel, Philippe Lemey, Ruth Tachezy, Sara Mostmans, Shin-Je Ghim, Koenraad Van Doorslaer, Melody Roelke, et al. 2007. "Ancient Papillomavirus-Host Co-Speciation in Felidae." *Genome Biology* 8 (4): R57. https://doi.org/10.1186/gb-2007-8-4-r57.

Revell, Liam J. 2012. "Phytools: An R Package for Phylogenetic Comparative Biology (and Other Things): *Phytools: R Package*." *Methods in Ecology and Evolution* 3 (2): 217–23. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

Richards, Rebecca M., Douglas R. Lowy, John T. Schiller, and Patricia M. Day. 2006. "Cleavage of the Papillomavirus Minor Capsid Protein, L2, at a Furin Consensus Site Is Necessary for Infection." *Proceedings of the National Academy of Sciences of the United States of America* 103 (5): 1522–27. https://doi.org/10.1073/pnas.0508815103.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Rose, Nathan R., and Robert J. Klose. 2014. "Understanding the Relationship between DNA Methylation and Histone Lysine Methylation." *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1839 (12): 1362–72. https://doi.org/10.1016/j.bbagrm.2014.02.007.

Sawyer, Sara L., Michael Emerman, and Harmit S. Malik. 2004. "Ancient Adaptive Evolution of the Primate Antiviral DNA-Editing Enzyme APOBEC3G." *PLoS Biology* 2 (9): E275. https://doi.org/10.1371/journal.pbio.0020275.

Schelhaas, Mario, Bhavin Shah, Michael Holzer, Peter Blattmann, Lena Kühling, Patricia M. Day, John T. Schiller, and Ari Helenius. 2012. "Entry of Human Papillomavirus Type 16 by Actin-Dependent, Clathrin- and Lipid Raft-Independent Endocytosis." *PLoS Pathogens* 8 (4): e1002657. https://doi.org/10.1371/journal.ppat.1002657.

Sen, George L., Jason A. Reuter, Daniel E. Webster, Lilly Zhu, and Paul A. Khavari. 2010. "DNMT1 Maintains Progenitor Function in Self-Renewing Somatic Tissue." *Nature* 463 (7280): 563–67. https://doi.org/10.1038/nature08683.

Sen, Goutam, Michael Flora, Gouri Chattopadhyay, Dennis M. Klinman, Andrew Lees, James J. Mond, and Clifford M. Snapper. 2004. "The Critical DNA Flanking Sequences of a CpG Oligodeoxynucleotide, but Not the 6 Base CpG Motif, Can Be Replaced with RNA without Quantitative or Qualitative Changes in Toll-like Receptor 9-Mediated Activity." *Cellular Immunology* 232 (1–2): 64–74. https://doi.org/10.1016/j.cellimm.2005.01.010.

Shahzad, Naveed, Masahiro Shuda, Tarik Gheit, Hyun Jin Kwun, Iris Cornet, Djamel Saidj, Claudia Zannetti, et al. 2013. "The T Antigen Locus of Merkel Cell Polyomavirus Downregulates Human Toll-like Receptor 9 Expression." *Journal of Virology* 87 (23): 13009–19. https://doi.org/10.1128/JVI.01786-13.

Sharp, Paul M., Therese M.F. Tuohy, and Krzysztof R. Mosurski. 1986. "Codon Usage in Yeast: Cluster Analysis Clearly Differentiates Highly and Lowly Expressed Genes." *Nucleic Acids Research* 14 (13): 5125–43. https://doi.org/10.1093/nar/14.13.5125.

Shen, Y.-Y., L. Liang, Z.-H. Zhu, W.-P. Zhou, D. M. Irwin, and Y.-P. Zhang. 2010. "Adaptive Evolution of Energy Metabolism Genes and the Origin of Flight in Bats." *Proceedings of the National Academy of Sciences* 107 (19): 8666–71. https://doi.org/10.1073/pnas.0912613107.

Smeele, Zoe E., Jennifer M. Burns, Koenraad Van Doorsaler, Rafaela S. Fontenele, Kara Waits, Daisy Stainton, Michelle R. Shero, et al. 2018. "Diverse Papillomaviruses Identified in Weddell Seals." *The Journal of General Virology*, February. https://doi.org/10.1099/jgv.0.001028.

Smith, Martin D., Joel O. Wertheim, Steven Weaver, Ben Murrell, Konrad Scheffler, and Sergei L. Kosakovsky Pond. 2015. "Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection." *Molecular Biology and Evolution* 32 (5): 1342–53. https://doi.org/10.1093/molbev/msv022.

Sorouri, Mahsa, Tyron Chang, Palmy Jesudhasan, Chelsea Pinkham, Nels C. Elde, and Dustin C. Hancks. 2020. "Signatures of Host–Pathogen Evolutionary Conflict Reveal MISTR—A Conserved MItochondrial STress Response Network." Edited by Mark L Siegal. *PLOS Biology* 18 (12): e3001045. https://doi.org/10.1371/journal.pbio.3001045.

Springer, M. S., E. C. Teeling, O. Madsen, M. J. Stanhope, and W. W. de Jong. 2001. "Integrated Fossil and Molecular Data Reconstruct Bat Echolocation." *Proceedings of the National Academy of Sciences* 98 (11): 6241–46. https://doi.org/10.1073/pnas.111551998.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1028 Stamatakis, Alexandros. 2014. “RAxML Version 8: A Tool for Phylogenetic Analysis and Post-
1029 Analysis of Large Phylogenies.” *Bioinformatics* 30 (9): 1312–13.
1030 https://doi.org/10.1093/bioinformatics/btu033.
1031 Steger, G, and S Corbach. 1997. “Dose-Dependent Regulation of the Early Promoter of Human
1032 Papillomavirus Type 18 by the Viral E2 Protein.” *Journal of Virology* 71 (1): 50–58.
1033 https://doi.org/10.1128/JVI.71.1.50-58.1997.
1034 Stephen F Altschul and Erickson Bruce W. 1985. “Significance of Nucleotide Sequence
1035 Alignments: A Method for Random Sequence Permutation That Preserves Dinucleotide
1036 and Codon Usage.” *Molecular Biology and Evolution*, November.
1037 https://doi.org/10.1093/oxfordjournals.molbev.a040370.
1038 Stepp, Wesley H., James D. Stamos, Simran Khurana, Alix Warburton, and Alison A. McBride.
1039 2017. “Sp100 Colocalizes with HPV Replication Foci and Restricts the Productive Stage
1040 of the Infectious Cycle.” Edited by Paul Francis Lambert. *PLOS Pathogens* 13 (10):
1041 e1006660. https://doi.org/10.1371/journal.ppat.1006660.
1042 Subudhi, Sonu, Noreen Rapin, and Vikram Misra. 2019. “Immune System Modulation and Viral
1043 Persistence in Bats: Understanding Viral Spillover.” *Viruses* 11 (2).
1044 https://doi.org/10.3390/v11020192.
1045 Takata, Matthew A., Daniel Gonçalves-Carneiro, Trinity M. Zang, Steven J. Soll, Ashley York,
1046 Daniel Blanco-Melo, and Paul D. Bieniasz. 2017. “CG Dinucleotide Suppression Enables
1047 Antiviral Defence Targeting Non-Self RNA.” *Nature* 550 (7674): 124–27.
1048 https://doi.org/10.1038/nature24039.
1049 Tan, Bing, Xing-Lou Yang, Xing-Yi Ge, Cheng Peng, Hai-Zhou Liu, Yun-Zhi Zhang, Li-Biao
1050 Zhang, and Zheng-Li Shi. 2017. “Novel Bat Adenoviruses with Low G+C Content Shed
1051 New Light on the Evolution of Adenoviruses.” *The Journal of General Virology* 98 (4):
1052 739–48. https://doi.org/10.1099/jgv.0.000739.
1053 Taubenberger, Jeffery K., and John C. Kash. 2010. “Influenza Virus Evolution, Host Adaptation,
1054 and Pandemic Formation.” *Cell Host & Microbe* 7 (6): 440–51.
1055 https://doi.org/10.1016/j.chom.2010.05.009.
1056 Teeling, E. C., O. Madsen, R. A. Van Den Bussche, W. W. de Jong, M. J. Stanhope, and M. S.
1057 Springer. 2002. “Microbat Paraphyly and the Convergent Evolution of a Key Innovation
1058 in Old World Rhinolophoid Microbats.” *Proceedings of the National Academy of*
1059 *Sciences* 99 (3): 1431–36. https://doi.org/10.1073/pnas.022477199.
1060 Thain, A., K. Webster, D. Emery, A. R. Clarke, and K. Gaston. 1997. “DNA Binding and
1061 Bending by the Human Papillomavirus Type 16 E2 Protein. Recognition of an Extended
1062 Binding Site.” *The Journal of Biological Chemistry* 272 (13): 8236–42.
1063 https://doi.org/10.1074/jbc.272.13.8236.
1064 Thierry, F., and M. Yaniv. 1987. “The BPV1-E2 Trans-Acting Protein Can Be Either an
1065 Activator or a Repressor of the HPV18 Regulatory Region.” *The EMBO Journal* 6 (11):
1066 3391–97.
1067 Thompson, J, and A Iwasaki. 2008. “Toll-like Receptors Regulation of Viral Infection and
1068 Disease☆.” *Advanced Drug Delivery Reviews* 60 (7): 786–94.
1069 https://doi.org/10.1016/j.addr.2007.11.003.
1070 Thompson, Mikayla R., John J. Kaminski, Evelyn A. Kurt-Jones, and Katherine A. Fitzgerald.
1071 2011. “Pattern Recognition Receptors and the Innate Immune Response to Viral
1072 Infection.” *Viruses* 3 (6): 920–40. https://doi.org/10.3390/v3060920.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1073 Tsagkogeorga, Georgia, Joe Parker, Elia Stupka, James A. Cotton, and Stephen J. Rossiter. 2013.
1074 "Phylogenomic Analyses Elucidate the Evolutionary Relationships of Bats." *Current*
1075 *Biology* 23 (22): 2262–67. https://doi.org/10.1016/j.cub.2013.09.014.
1076 Uhlorn, Brittany L., Eduardo R. Gamez, Shuaizhi Li, and Samuel K. Campos. 2020.
1077 "Attenuation of CGAS/STING Activity during Mitosis." *Life Science Alliance* 3 (9).
1078 https://doi.org/10.26508/lsa.201900636.
1079 Uhlorn, Brittany L., Robert Jackson, Shuaizhi Li, Shauna M. Bratton, Koenraad Van Doorslaer,
1080 and Samuel K. Campos. 2020. "Vesicular Trafficking Permits Evasion of CGAS/STING
1081 Surveillance during Initial Human Papillomavirus Infection." Edited by Paul Francis
1082 Lambert. *PLOS Pathogens* 16 (11): e1009028.
1083 https://doi.org/10.1371/journal.ppat.1009028.
1084 Upadhyay, Mohita, and Perumal Vivekanandan. 2015. "Depletion of CpG Dinucleotides in
1085 Papillomaviruses and Polyomaviruses: A Role for Divergent Evolutionary Pressures."
1086 Edited by Robert D. Burk. *PLOS ONE* 10 (11): e0142368.
1087 https://doi.org/10.1371/journal.pone.0142368.
1088 Van Doorslaer, Koenraad. 2013. "Evolution of the Papillomaviridae." *Virology* 445 (1–2): 11–
1089 20. https://doi.org/10.1016/j.virol.2013.05.012.
1090 Van Doorslaer, Koenraad, Zigui Chen, Hans-Ulrich Bernard, Paul K. S. Chan, Rob DeSalle,
1091 Joakim Dillner, Ola Forslund, et al. 2018. "ICTV Virus Taxonomy Profile:
1092 Papillomaviridae." *The Journal of General Virology* 99 (8): 989–90.
1093 https://doi.org/10.1099/jgv.0.001105.
1094 Van Doorslaer, Koenraad, and Joakim Dillner. 2019. "The Launch of an International Animal
1095 Papillomavirus Reference Center." *Viruses* 11 (1). https://doi.org/10.3390/v11010055.
1096 Van Doorslaer, Koenraad, Zhiwen Li, Sandhya Xirasagar, Piet Maes, David Kaminsky, David
1097 Liou, Qiang Sun, Ramandeep Kaur, Yentram Huyen, and Alison A. McBride. 2017. "The
1098 Papillomavirus Episteme: A Major Update to the Papillomavirus Sequence Database."
1099 *Nucleic Acids Research* 45 (D1): D499–506. https://doi.org/10.1093/nar/gkw879.
1100 Van Doorslaer, Koenraad, and Alison A. McBride. 2016. "Molecular Archeological Evidence in
1101 Support of the Repeated Loss of a Papillomavirus Gene." *Scientific Reports* 6
1102 (September): 33028. https://doi.org/10.1038/srep33028.
1103 Van Doorslaer, Koenraad, Qina Tan, Sandhya Xirasagar, Sandya Bandaru, Vivek Gopalan,
1104 Yasmin Mohamoud, Yentram Huyen, and Alison A. McBride. 2013. "The
1105 Papillomavirus Episteme: A Central Resource for Papillomavirus Sequence Data and
1106 Analysis." *Nucleic Acids Research* 41 (Database issue): D571-578.
1107 https://doi.org/10.1093/nar/gks984.
1108 Villiers, Ethel-Michele de, Claude Fauquet, Thomas R. Broker, Hans-Ulrich Bernard, and Harald
1109 zur Hausen. 2004. "Classification of Papillomaviruses." *Virology* 324 (1): 17–27.
1110 https://doi.org/10.1016/j.virol.2004.03.033.
1111 Vincent, Isabelle E., Claudia Zannetti, Julie Lucifora, Helene Norder, Ulrike Protzer, Pierre
1112 Hainaut, Fabien Zoulim, et al. 2011. "Hepatitis B Virus Impairs TLR9 Expression and
1113 Function in Plasmacytoid Dendritic Cells." *PloS One* 6 (10): e26315.
1114 https://doi.org/10.1371/journal.pone.0026315.
1115 Vinokurova, Svetlana, and Magnus von Knebel Doeberitz. 2011. "Differential Methylation of
1116 the HPV 16 Upstream Regulatory Region during Epithelial Differentiation and
1117 Neoplastic Transformation." Edited by Torbjorn Ramqvist. *PLoS ONE* 6 (9): e24451.
1118 https://doi.org/10.1371/journal.pone.0024451.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1119 Wacharapluesadee, Supaporn, Chee Wah Tan, Patarapol Maneeorn, Prateep Duengkae, Feng
1120 Zhu, Yutthana Joyjinda, Thongchai Kaewpom, et al. 2021. "Evidence for SARS-CoV-2
1121 Related Coronaviruses Circulating in Bats and Pangolins in Southeast Asia." *Nature*
1122 *Communications* 12 (1): 972. https://doi.org/10.1038/s41467-021-21240-1.
1123 Wang, Joshua W., and Richard B.S. Roden. 2013. "L2, the Minor Capsid Protein of
1124 Papillomavirus." *Virology* 445 (0): 175–86. https://doi.org/10.1016/j.virol.2013.04.017.
1125 Wang, Li-Gen, Tommy Tsan-Yuk Lam, Shuangbin Xu, Zehan Dai, Lang Zhou, Tingze Feng,
1126 Pingfan Guo, et al. 2020. "Treeio: An R Package for Phylogenetic Tree Input and Output
1127 with Richly Annotated and Associated Data." *Molecular Biology and Evolution* 37 (2):
1128 599–603. https://doi.org/10.1093/molbev/msz240.
1129 Warren, Cody J., Koenraad Van Doorslaer, Ahwan Pandey, Joaquin M. Espinosa, and Dohun
1130 Pyeon. 2015. "Role of the Host Restriction Factor APOBEC3 on Papillomavirus
1131 Evolution." *Virus Evolution* 1 (1). https://doi.org/10.1093/ve/vev015.
1132 Warren, Cody J., Tao Xu, Kejun Guo, Laura M. Griffin, Joseph A. Westrich, Denis Lee, Paul F.
1133 Lambert, Mario L. Santiago, and Dohun Pyeon. 2015. "APOBEC3A Functions as a
1134 Restriction Factor of Human Papillomavirus." *Journal of Virology* 89 (1): 688–702.
1135 https://doi.org/10.1128/JVI.02383-14.
1136 Weber, Friedemann, Valentina Wagner, Simon B. Rasmussen, Rune Hartmann, and Søren R.
1137 Paludan. 2006. "Double-Stranded RNA Is Produced by Positive-Strand RNA Viruses and
1138 DNA Viruses but Not in Detectable Amounts by Negative-Strand RNA Viruses." *Journal*
1139 *of Virology* 80 (10): 5059–64. https://doi.org/10.1128/JVI.80.10.5059-5064.2006.
1140 Weber, Michael, Ines Hellmann, Michael B Stadler, Liliana Ramos, Svante Pääbo, Michael
1141 Rebhan, and Dirk Schübeler. 2007. "Distribution, Silencing Potential and Evolutionary
1142 Impact of Promoter DNA Methylation in the Human Genome." *Nature Genetics* 39 (4):
1143 457–66. https://doi.org/10.1038/ng1990.
1144 Wertheim, Joel O., Ben Murrell, Martin D. Smith, Sergei L. Kosakovsky Pond, and Konrad
1145 Scheffler. 2015. "RELAX: Detecting Relaxed Selection in a Phylogenetic Framework."
1146 *Molecular Biology and Evolution* 32 (3): 820–32.
1147 https://doi.org/10.1093/molbev/msu400.
1148 Willemsen, Anouk, and Ignacio G. Bravo. 2019. "Origin and Evolution of Papillomavirus
1149 (Onco)Genes and Genomes." *Philosophical Transactions of the Royal Society of London.*
1150 *Series B, Biological Sciences* 374 (1773): 20180303.
1151 https://doi.org/10.1098/rstb.2018.0303.
1152 Wong, Emily HM, David K Smith, Raul Rabadan, Malik Peiris, and Leo LM Poon. 2010.
1153 "Codon Usage Bias and the Evolution of Influenza A Viruses. Codon Usage Biases of
1154 Influenza Virus." *BMC Evolutionary Biology* 10 (1): 253. https://doi.org/10.1186/1471-
1155 2148-10-253.
1156 Xie, Jiazheng, Yang Li, Xurui Shen, Geraldine Goh, Yan Zhu, Jie Cui, Lin-Fa Wang, Zheng-Li
1157 Shi, and Peng Zhou. 2018. "Dampened STING-Dependent Interferon Activation in Bats."
1158 *Cell Host & Microbe* 23 (3): 297-301.e4. https://doi.org/10.1016/j.chom.2018.01.006.
1159 Yi, A. K., M. Chang, D. W. Peckham, A. M. Krieg, and R. F. Ashman. 1998. "CpG
1160 Oligodeoxyribonucleotides Rescue Mature Spleen B Cells from Spontaneous Apoptosis
1161 and Promote Cell Cycle Entry." *Journal of Immunology (Baltimore, Md.: 1950)* 160 (12):
1162 5898–5906.

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1163 Yu, Guangchuang, Tommy Tsan-Yuk Lam, Huachen Zhu, and Yi Guan. 2018. "Two Methods
1164 for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree." *Molecular*
1165 *Biology and Evolution* 35 (12): 3041–43. https://doi.org/10.1093/molbev/msy194.
1166 Zhang, Guojie, Christopher Cowled, Zhengli Shi, Zhiyong Huang, Kimberly A. Bishop-Lilly,
1167 Xiaodong Fang, James W. Wynne, et al. 2013. "Comparative Analysis of Bat Genomes
1168 Provides Insight into the Evolution of Flight and Immunity." *Science (New York, N.Y.)*
1169 339 (6118): 456–60. https://doi.org/10.1126/science.1230835.
1170 Zucchini, Nicolas, Gilles Bessou, Stephanie Traub, Scott H. Robbins, Satoshi Uematsu, Shizuo
1171 Akira, Lena Alexopoulou, and Marc Dalod. 2008. "Cutting Edge: Overlapping Functions
1172 of TLR7 and TLR9 for Innate Defense against a Herpesvirus Infection." *The Journal of*
1173 *Immunology* 180 (9): 5799–5803. https://doi.org/10.4049/jimmunol.180.9.5799.
1174

1175

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure Legends

Figure 1 Evolutionary relationship of novel bat papillomaviruses

(A) Maximum-likelihood phylogenetic tree inferred using concatenated E1, E2, and L1 protein sequences. Papillomaviruses associated with *Chiroptera* are highlighted Yangochiroptera (orange) and Yinpterochiroptera (red). Papillomavirus genera are collapsed (number of types within each genus are indicated in parentheses). Bootstrap generated branch support values are given using symbols and color gradient. Host species are indicated using Sonoran Desert dwelling animals. The red arrow indicates the subtree used for further analyses throughout the manuscript.

(B) Pairwise identity plot with percentage pairwise identities provided in colored boxes for the L1 nucleotide sequences.

Figure 2 Co-evolution of papillomaviruses

(A) Optimized tanglegram between subtree based on concatenated E1–E2–L1 maximum-likelihood phylogenetic tree (see **Figure 1**) and associated host species. Host species tree was downloaded from www.timetree.org. Papillomaviruses are linked to their host phylogenies. Papillomaviruses associated with Chiroptera are highlighted Yangochiroptera (orange) and Yinpterochiroptera (blue).

(B) Procrustean Approach to Cophylogeny analysis based on the interaction network and phylogenies shown in (A) supports that papillomaviruses coevolved with their hosts. The observed best-fit Procrustean super-imposition (red dotted line line) lies outside of the 95% confidence interval (shaded area of the curves) of the distribution of network randomizations in the null model.

(C) As in (B) but using a subset of the interaction network and phylogenies (indicated by red arrow in (A))

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 3 CpG dinucleotide sequences are significantly depleted in papillomavirus genomes

The observed vs. expected (O/E) ratios of each dinucleotide in the papillomavirus genomes sequences shown in **Figure 2** were calculated using a custom wrapper around the CompSeq program from the EMBOSS software suite. The red line indicates that the sequence is seen as often as would be expected by chance.

Figure 4 CpG content is significantly lower in papillomaviruses associated with Yangochiroptera compared to related viruses

(A) A maximum likelihood phylogenetic tree is shown comparing the O/E ratios of CpG dinucleotides. Viruses infecting Yangochiroptera (red), Yinpterochiroptera (green), and related hosts (grey) are indicated.

(B) Mean (+/- standard deviation) CpG observed vs. expected (O/E) ratios for each group of viruses are compared using a one-way ANOVA with Tukey's posthoc test.

(C) CpG observed vs. expected (O/E) ratios are compared to total GC content for each viral genome in (A).

Figure 5 Yangochiroptera papillomaviruses have a restricted codon usage

(A) Codon usage tables for each virus in Figure 2 were compared using the 'codcmp' program from the EMBOSS software suite. Root-mean-square deviation (RMSD) values for each pairwise comparison are plotted as Box-and-whisker plots with the outliers (colored circles) identified using Tukey's method. Individual values are shown as a single black dot.

(B) Amino-acid composition was calculated as described in materials and methods. Mean values +/- standard deviation is plotted.

(C) RSCU values for the indicated amino acid/codons were calculated and plotted as Box-and-whisker plots with the outliers (colored circles) identified using Tukey's method. RSCU

bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

values for each amino acid were compared using a two-way ANOVA with Tukey's posthoc test. Significance is indicated as shown in the legend.

Figure 6 Yangochiroptera TLR9 is evolving under diversifying selection

(A) Structure of horse TLR9 in complex with agonistic DNA (PDB: 3WPC) (Ohto et al. 2015). Amino acids of interest are highlighted.

(B) Maximum likelihood phylogenetic tree of mammalian TLR9 sequences clusters Yangochiroptera and Yinpterochiroptera separate from the mammalian TLR9. Red branches display evidence of episodic diversifying selection as identified by aBRSEL (Smith et al. 2015). Alignments show sequences of interest. The sequence logo is based on the alignment of 29 non-chiropteran TLR9 sequences. Numbering is based on the mouse TLR9. Residues indicated with $ were identified as being selected using FEL. Residues highlighted with * were previously identified as evolving under diversifying selection (Escalera-Zamudio et al. 2015), while residues with # were shown to be functionally important through site directed mutagenesis (Ohto et al. 2015.

Figure 7 Yangochiroptera papillomaviruses depleted a TLR9 recognition motif from their genomes

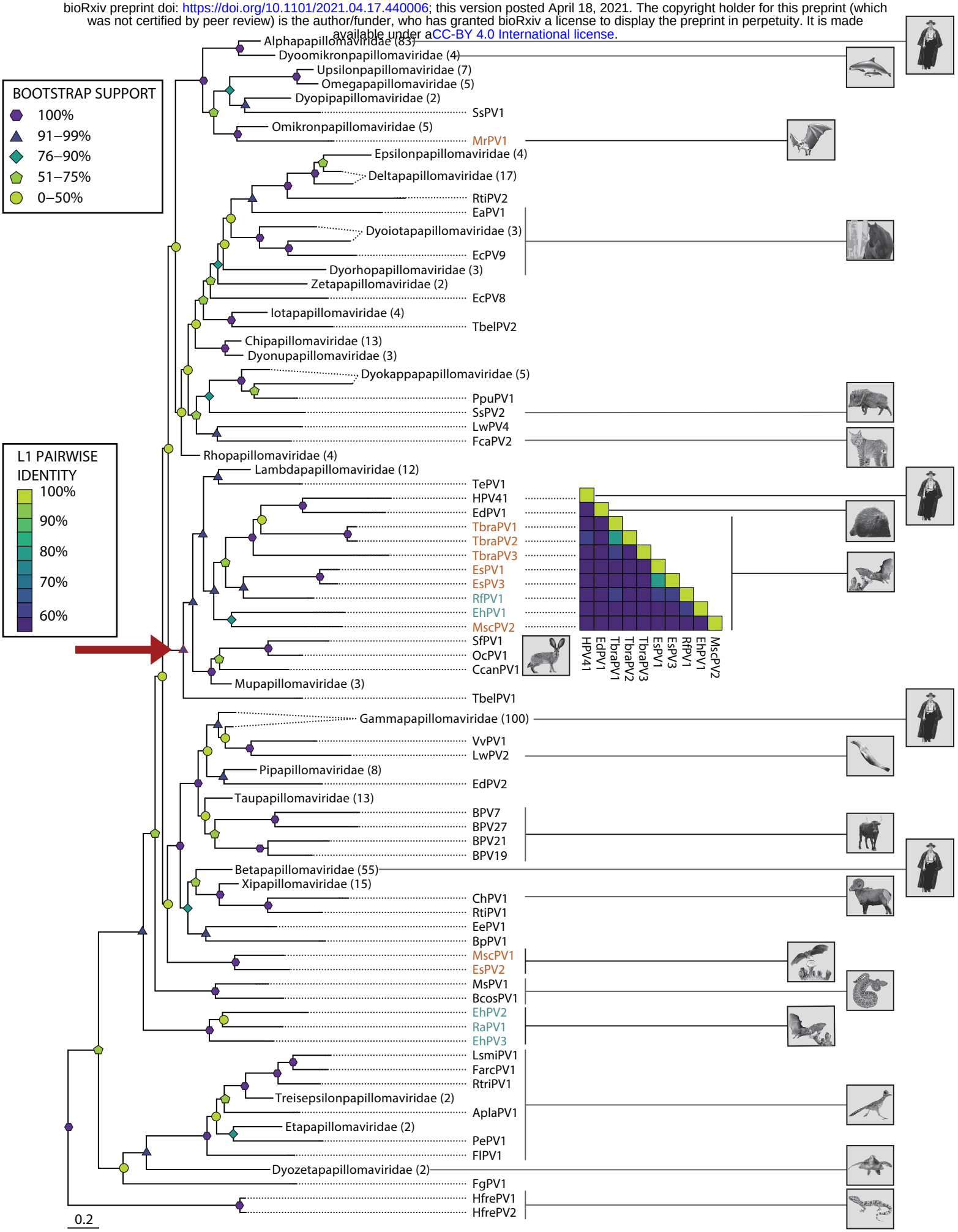
(A) The observed vs. expected (O/E) ratios of each N-CG-N tetramer in the Yangochiroptera papillomavirus genomes sequences were calculated using a custom wrapper around the CompSeq program from the EMBOSS software suite. Mean values +/- standard deviation are plotted.

(B) The observed vs. expected (O/E) ratios of each N-CG-N tetramer in the different groups was calculated as in A. The proportion of these ratios are shown to provide a normalized view of tetramer depletion across papillomavirus genomes shown in Figure 2.

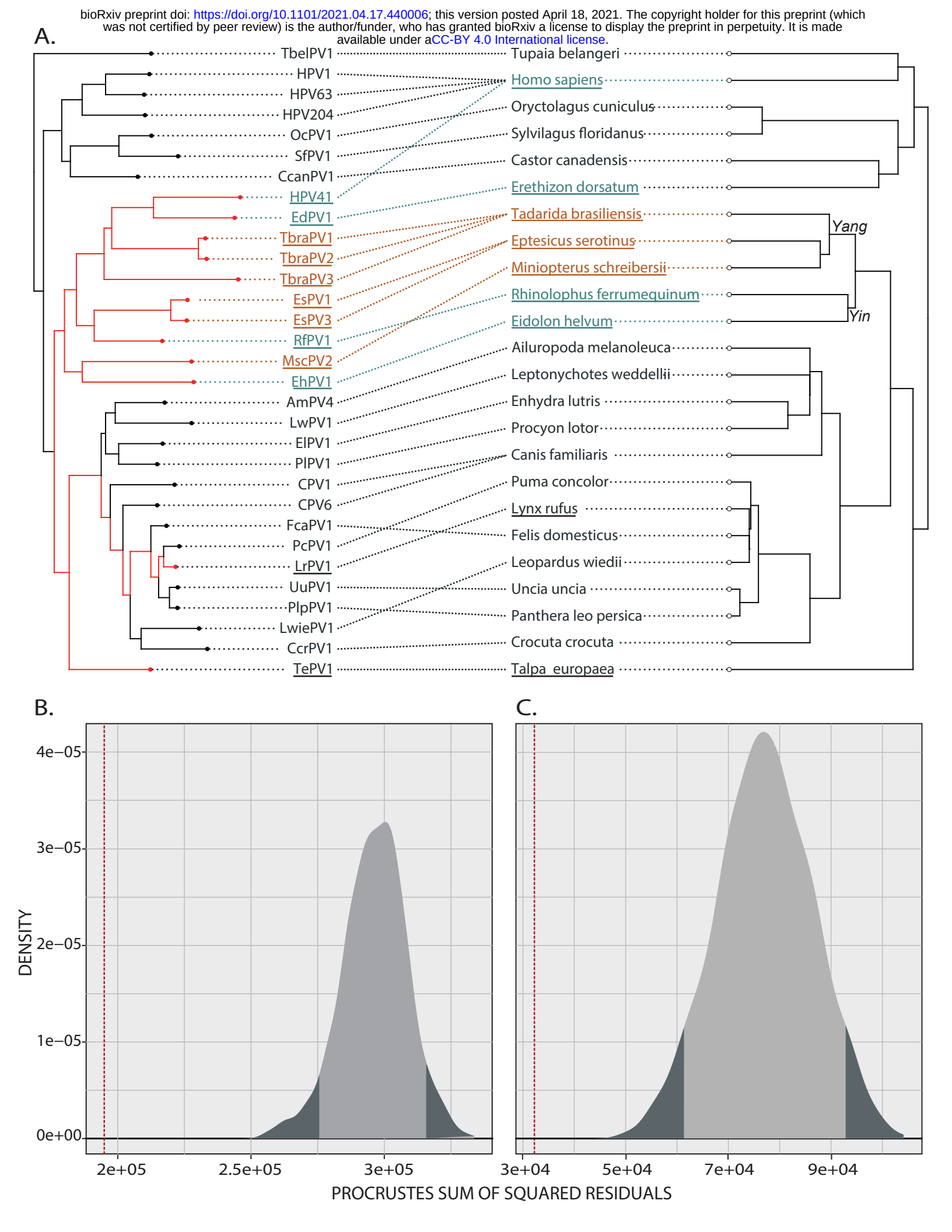
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1253 **(C)** The Yang vs. Yin N-CG-N proportion (as in B) are plotted as brown dots and compared to
1254 1000 randomly shuffled sequences (green violin) plots. Only ACGT is statistically
1255 underrepresented in the Yangochiroptera.
1256
1257

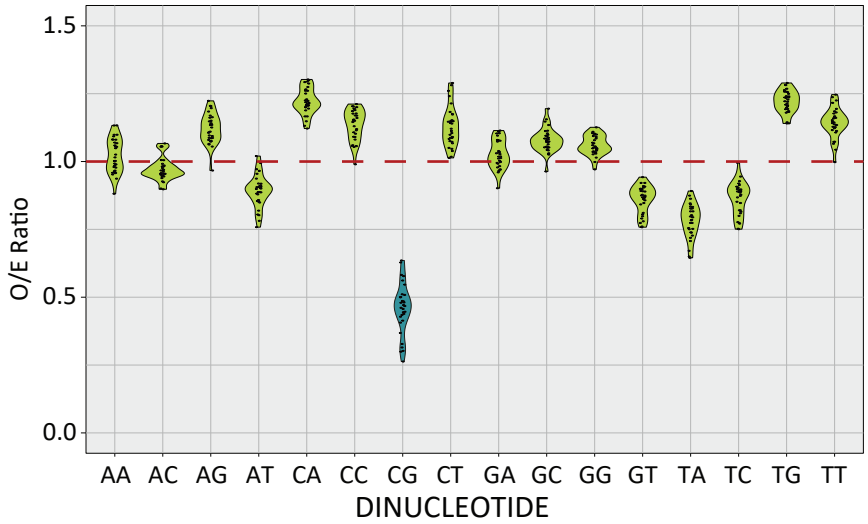
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



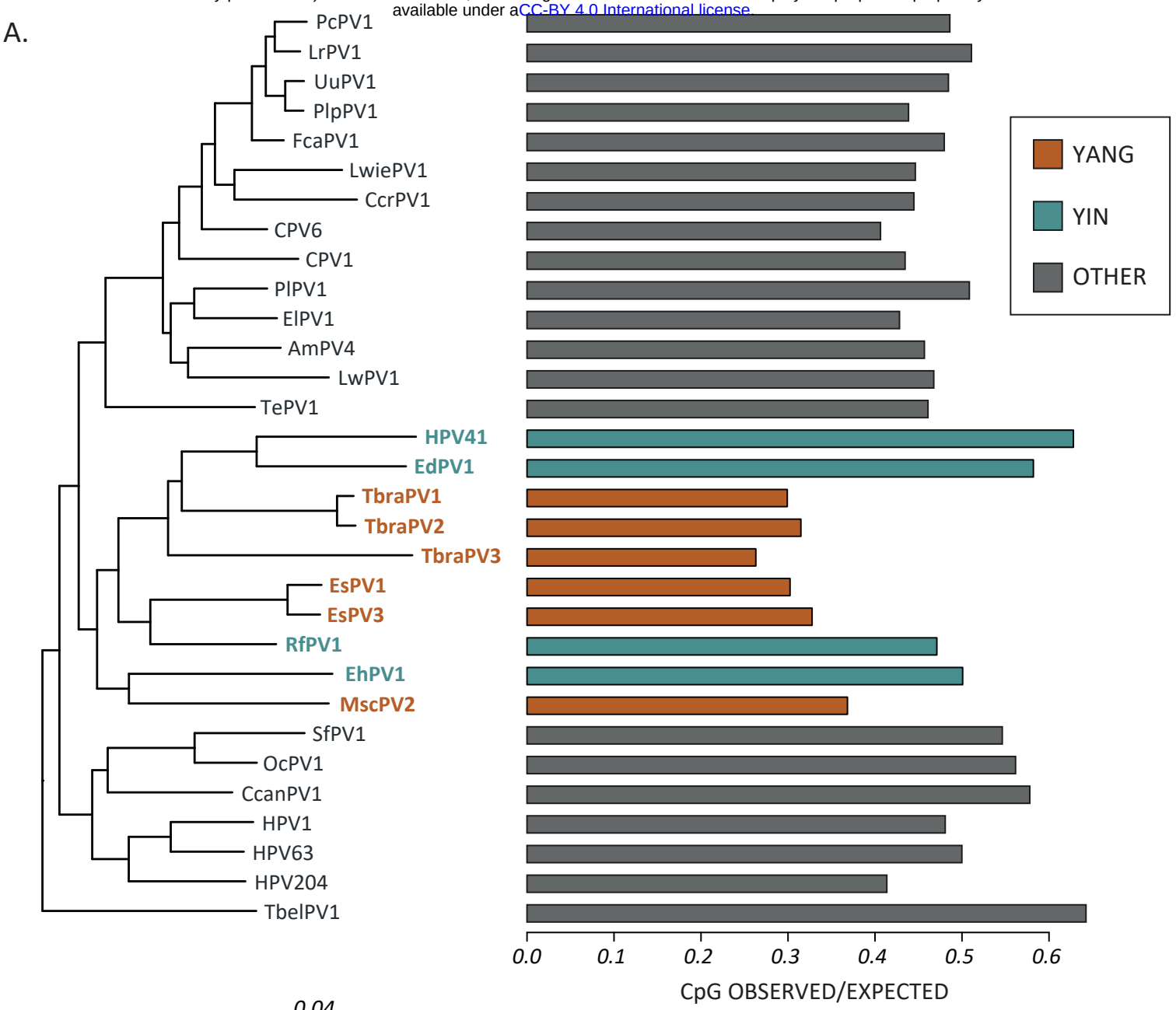
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

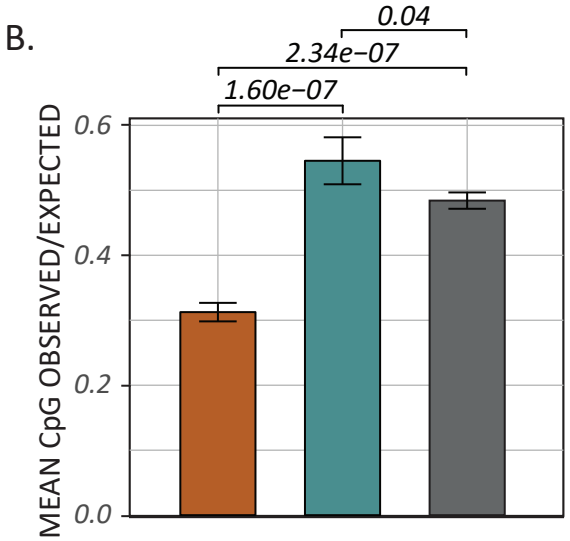


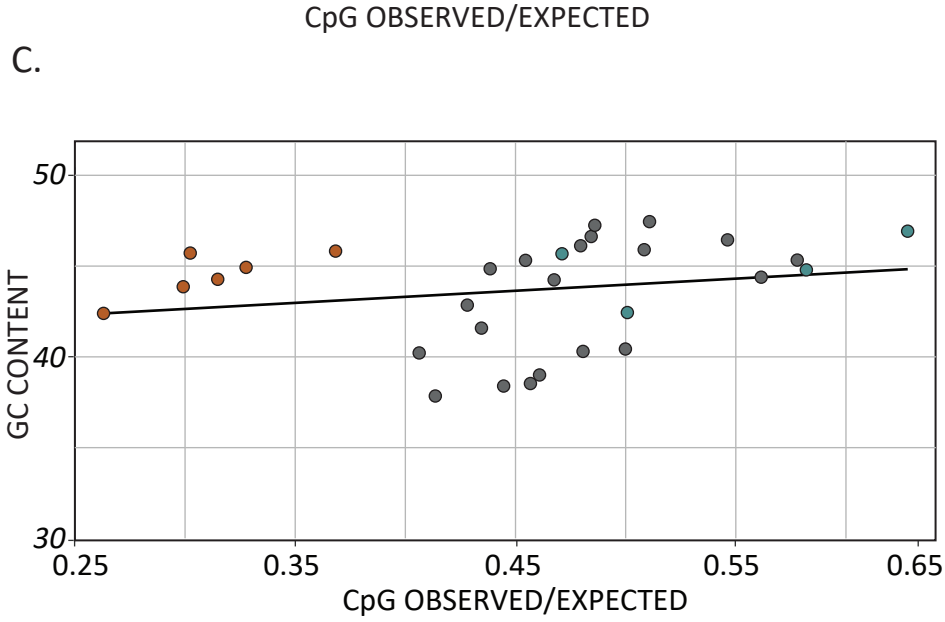
bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

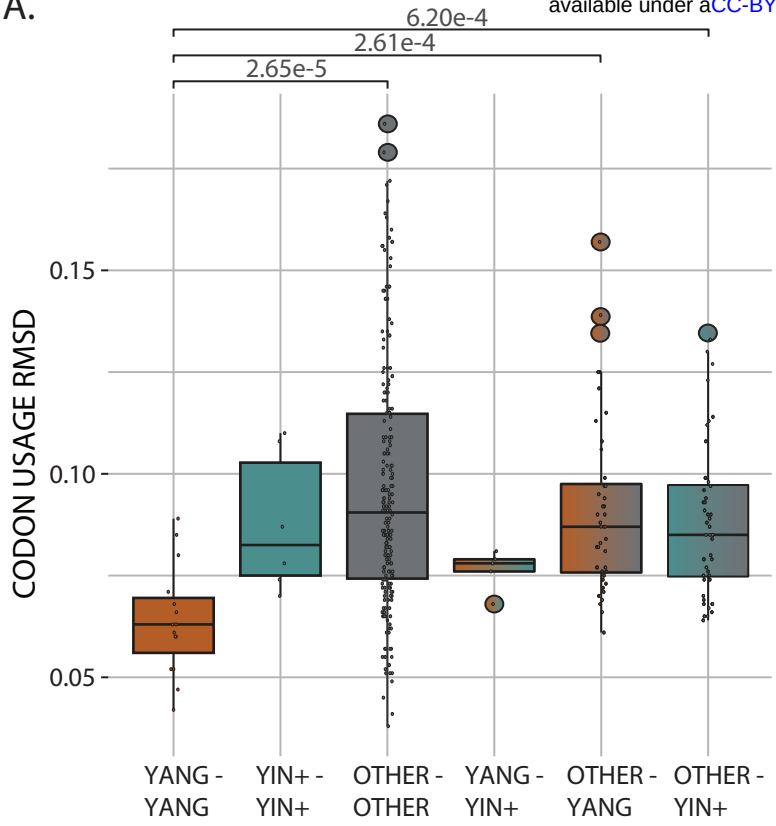




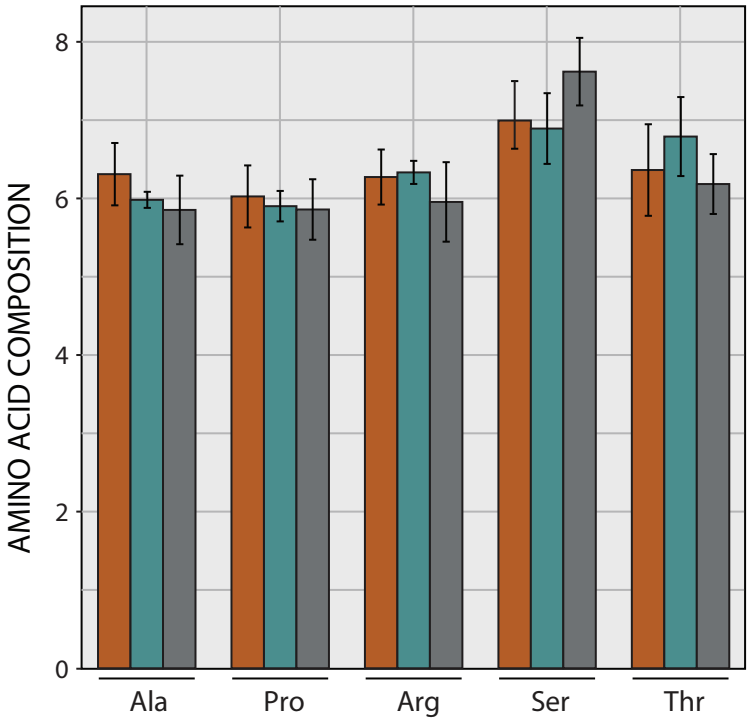


bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A.

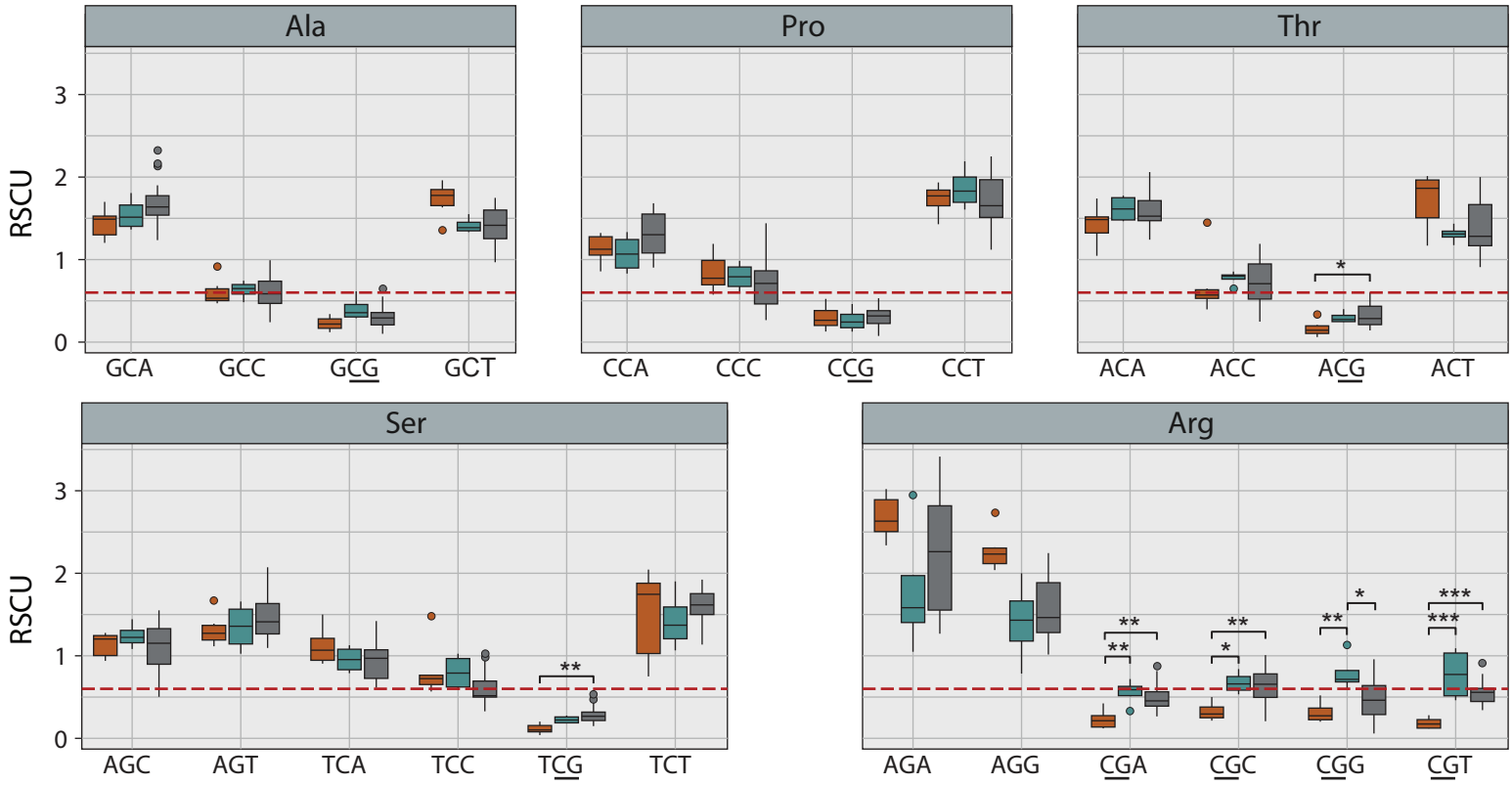


B.



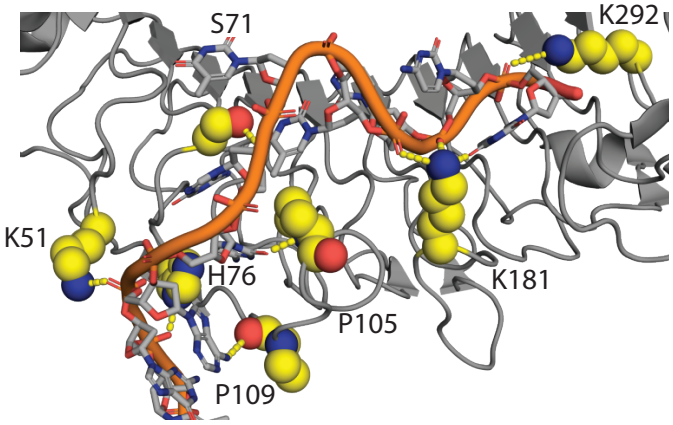
C.

*	**	***
<0.05	<0.01	<0.001

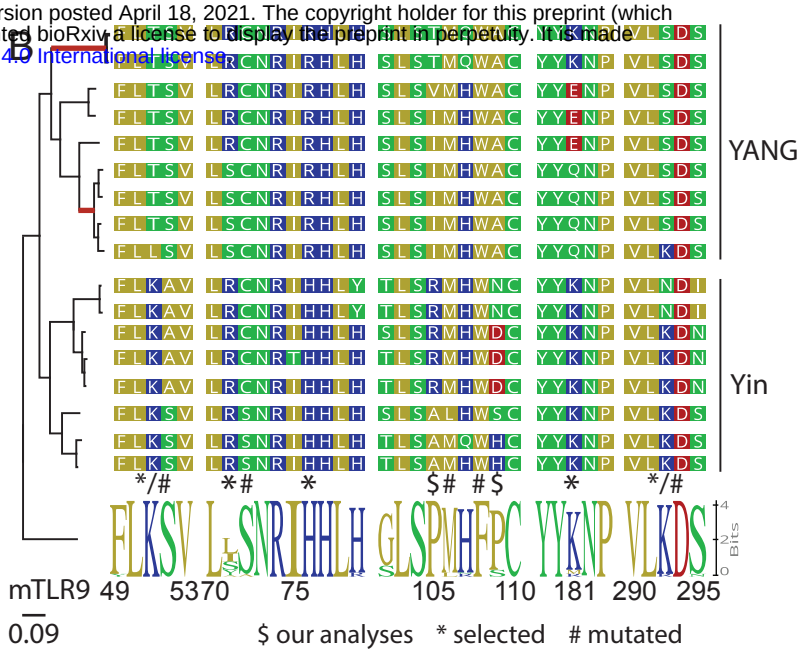


bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A



B



bioRxiv preprint doi: https://doi.org/10.1101/2021.04.17.440006; this version posted April 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.