

RESEARCH ARTICLE

# Niche adaptation and viral transmission of human papillomaviruses from archaic hominins to modern humans

Zigui Chen<sup>1\*</sup>, Rob DeSalle<sup>2</sup>, Mark Schiffman<sup>3</sup>, Rolando Herrero<sup>4,5</sup>, Charles E. Wood<sup>6a</sup>, Julio C. Ruiz<sup>7</sup>, Gary M. Clifford<sup>8</sup>, Paul K. S. Chan<sup>1</sup>, Robert D. Burk<sup>8\*</sup>

**1** Departments of Microbiology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China, **2** Sackler Institute of Comparative Genomics, American Museum of Natural History, New York, NY, United States of America, **3** Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, United States of America, **4** International Agency for Research on Cancer, World Health Organization, Lyon, France, **5** Proyecto Epidemiológico Guanacaste, Fundación INCIENSA, San José, Costa Rica, **6** Department of Pathology, Wake Forest School of Medicine, Winston-Salem, NC, United States of America, **7** Department of Veterinary Sciences, The University of Texas MD Anderson Cancer Center, Bastrop, Texas, United States of America, **8** Departments of Pediatrics, Microbiology and Immunology; Epidemiology and Population Health; Obstetrics, Gynecology and Woman's Health, Albert Einstein College of Medicine, Bronx, NY, United States of America

\* Current address: Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT, United States of America  
\* [zigui.chen@cuhk.edu.hk](mailto:zigui.chen@cuhk.edu.hk) (ZC); [robert.burk@einstein.yu.edu](mailto:robert.burk@einstein.yu.edu) (RDB)



OPEN ACCESS

**Citation:** Chen Z, DeSalle R, Schiffman M, Herrero R, Wood CE, Ruiz JC, et al. (2018) Niche adaptation and viral transmission of human papillomaviruses from archaic hominins to modern humans. *PLoS Pathog* 14(11): e1007352. <https://doi.org/10.1371/journal.ppat.1007352>

**Editor:** Nels C. Elde, University of Utah, UNITED STATES

**Received:** June 28, 2018

**Accepted:** September 22, 2018

**Published:** November 1, 2018

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files except for the sequences of SscPV1/2/3 and MmpV2/3/4 which are available from Genbank under the accession numbers of JF304765 to JF304767 and MG837557 to MG837559, respectively.

**Funding:** This work was supported in part by the National Cancer Institute (CA78527) (RDB), the Einstein-Rockefeller-CUNY Center for AIDS funded by the NIH (P30 AI-124414) (RDB), and the

## Abstract

Recent discoveries on the origins of modern humans from multiple archaic hominin populations and the diversity of human papillomaviruses (HPVs) suggest a complex scenario of virus-host evolution. To evaluate the origin of HPV pathogenesis, we estimated the phylogeny, timing, and dispersal of HPV16 variants using a Bayesian Markov Chain Monte Carlo framework. To increase precision, we identified and characterized non-human primate papillomaviruses from New and Old World monkeys to set molecular clock models. We demonstrate specific host niche adaptation of primate papillomaviruses with subsequent coevolution with their primate hosts for at least 40 million years. Analyses of 212 HPV16 complete genomes and 3582 partial sequences estimated ancient divergence of HPV16 variants (between A and BCD lineages) from their most recent common ancestors around half a million years ago, roughly coinciding with the timing of the split between archaic Neanderthals and modern *Homo sapiens*, and nearly three times longer than divergence times of modern *Homo sapiens*. HPV16 A lineage variants were significantly underrepresented in present African populations, whereas the A sublineages were highly prevalent in European (A1-3) and Asian (A4) populations, indicative of viral sexual transmission from Neanderthals to modern non-African humans through multiple interbreeding events in the past 80 thousand years. Remarkably, the human leukocyte antigen B\*07:02 and C\*07:02 alleles associated with increased risk in cervix cancer represent introgressed regions from Neanderthals in present-day Eurasians. The archaic *hominin-host-switch* model was also supported by other HPV variants. Niche adaptation and virus-host codivergence appear to influence the pathogenesis of papillomaviruses.

Einstein Cancer Research Center (P30CA013330) from the National Cancer Institute (RDB). RD acknowledges the Korein Foundation and the Sackler Institute for Comparative Genomics at the AMNH for their continued support. ZC was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 24100716). Work at IARC was supported by a grant from the Institut National du Cancer (INCa), France (SHSESP 16-006) (GMC). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

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

## Author summary

Epidemiologic studies have demonstrated that persistent infection of select oncogenic human papillomaviruses (HPVs) is the main cause of cervix precancer and cancer. Nevertheless, our knowledge of the underlying evolutionary mechanisms driving the divergence and emergence of viral oncogenicity in specific types of HPVs is incomplete. To better understand the molecular evolution of oncogenic HPVs, we isolated viruses from non-human primates, evaluated papillomavirus molecular clock models, and estimated the divergence times of HPV16 and other HPV type variants from their most recent common ancestors. Primate PV-host tissue tropisms indicated niche adaptation of viruses to host ecosystems as the first stage of the evolution of oncogenic HPVs. The data also provided evidence of ancient codivergence of HPV variants with archaic hominins and recent viral transmission from Neanderthals to modern non-African humans through sexual intercourse. Understanding the evolution of papillomaviruses should provide important biological insights and suggest mechanisms underlying HPV-induced cervical cancer, since niche adaptation rather than oncogenicity drives viral fitness.

## Introduction

Papillomaviruses (PVs) are ubiquitous, non-enveloped, small double-stranded circular DNA viruses that cause proliferation of epithelial cells in a wide range of vertebrate host species, from reptiles to mammals [1, 2]. Currently, over 200 PVs infecting primate hosts (human and non-human) have been characterized and shown to group predominantly within 3 highly divergent genera—*Alphapapillomavirus*, *Betapapillomavirus*, and *Gammmapapillomavirus* [3]. All oncogenic PVs associated with the development of cervical carcinoma, including human PV (HPV) types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59 and *Macaca fascicularis* PV type 3 (MfPV3), share a common ancestor within the *Alphapapillomavirus* [4–7]. Among these oncogenic types, which are sexually transmitted primarily through intercourse [8, 9], HPV16 is globally the most prevalent HPV type detected, suggesting an increased fitness [10–12]. Moreover, HPV16 is also the most common HPV type detected in cervical cancer, which is the fourth most common cancer among women worldwide [13]. Nevertheless, most exposures to HPV types are transient, and many PVs appear to be more commensal than pathogenic [14].

Strict coevolution of a host and its pathogen is more likely if the pathogen is transmitted vertically and there is little or no cross-species acquisition. Persistent infection by pathogens generally indicates that they are well adapted to their host and that extinction will be rare so long as the host survives. Hence, in scenarios of coevolution, the evolutionary history of a pathogen should mirror that of its host, both in divergence times and phylogenetic history (Fahrenholz's rule) [15, 16]. These criteria have been shown to hold for feline PVs within the genus *Lambdapapillomavirus* isolated from oral lesions [17]. On the other hand, horizontal transmission of pathogens through host switching without restricted species specificity will produce a very different evolutionary history between host and pathogen. In hosts harboring many different types of PVs (e.g., bovines, humans, and macaques), the selection pressure exerted by PVs on their hosts appears negligible in comparison with what the hosts exert on the PV pathogens. Within human populations, for example, the ancient dispersal of HPV variants (e.g., HPV16 and HPV58) challenges a simple evolutionary pattern of viruses migrating with modern *Homo sapiens* [18], and instead indicates codivergence of viruses with archaic hominins

and transmission to modern humans [19, 20]. The genetic heterogeneity of PVs implies a complex evolutionary history with many interacting factors, including but not limited to virus-host codivergence, tissue tropism, lineage sorting, transmission, recombination, and natural selection [21, 22]. Understanding the capacity for, and history of, viral adaptation to host ecological environments is essential for understanding the genetic basis of HPV carcinogenicity [23]. However, the origin and evolution of oncogenic PVs remains poorly understood.

In this report, we estimate the divergence times of HPV16 and other oncogenic HPV types using a well-established Bayesian molecular clock model with newly characterized primate PV genomes that validate the divergence times of primate HPVs within niche-specific clades. Our analyses of the evolutionary dynamics of primate PVs, including specific focus on HPV16 variants, provide novel insights into the complex phylodynamic interactions between viruses and hosts and their pathologic outcomes.

## Results

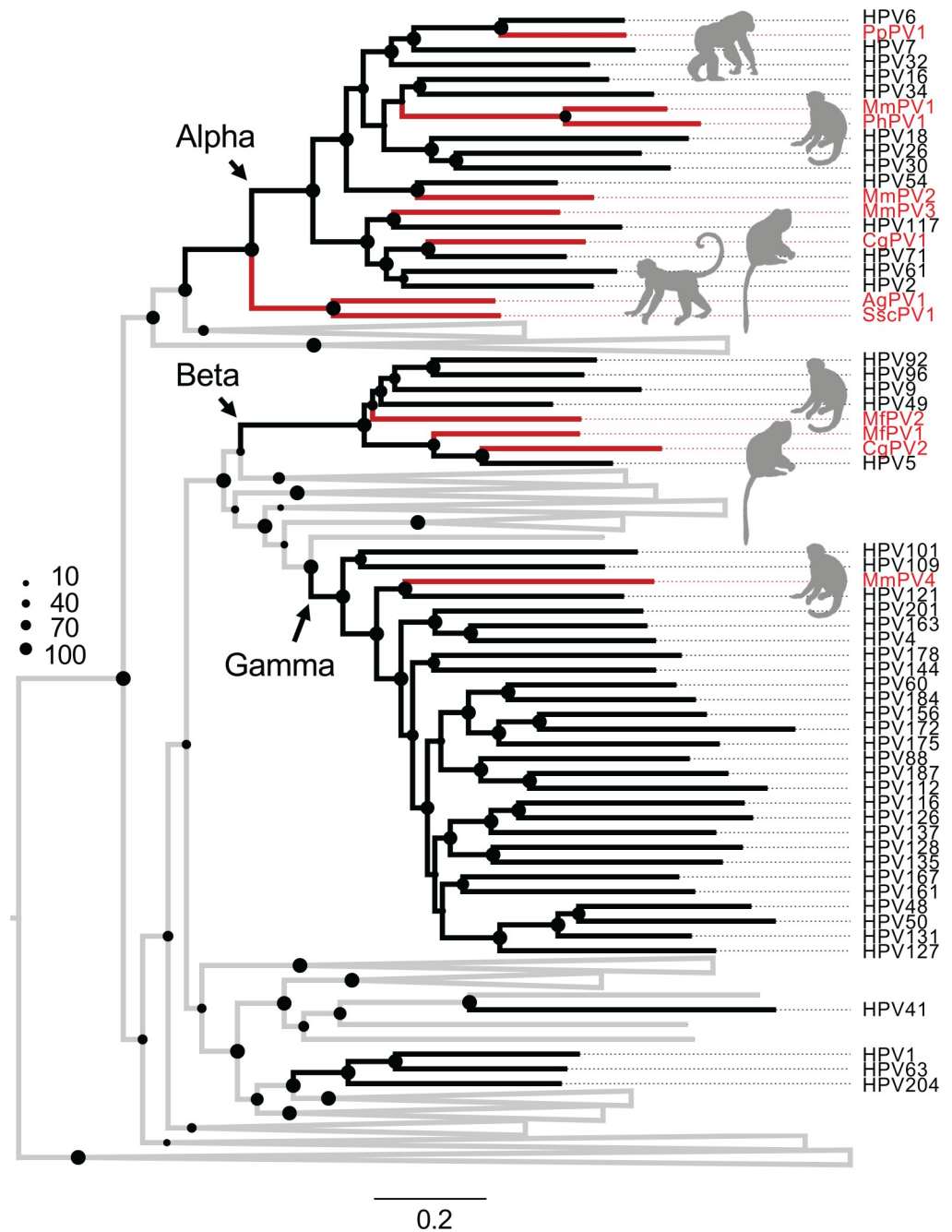
### Genomic characterization of novel non-human primate papillomaviruses

In an effort to study the diversity of non-human primate PVs (NHP-PVs) to better understand the evolution of oncogenic HPVs, we screened cervicovaginal specimens from 10 adult female squirrel monkeys (*Saimiri sciureus*), and the paired oral, perianal, and genital samples from 8 adult rhesus monkeys (*Macaca mulatta*) (4 females and 4 males). Three novel *Saimiri sciureus* PV types (SscPV1, 2 and 3) and three novel *Macaca mulatta* PV types (MmPV2, 3 and 4) were isolated and characterized and had genomes ranging in size from 7424 bp to 8051 bp (S1 Table). All genomes contained five early genes (E6, E7, E1, E2, and E4), two late genes (L2 and L1), and an upstream regulatory region (URR) between L1 and E6 genes. Phylogenetic trees based on the nucleotide sequence alignment of the concatenated four open reading frames (ORFs) (E1, E2, L2, and L1) (Fig 1 and S1 Fig) or individual genes, e. g., E1 or L1 ORFs (S2 Fig, S3 Fig and S4 Fig) support a monophyletic clade grouping SscPV1/2/3 and howler monkey *Alouatta guariba* PV 1 (AgPV1, KP861980) [24] within the genus *Dyoomikronpapillomavirus*. MmPV2 and MmPV3 cluster into the genus *Alphapapillomavirus*, with the closest HPVs being HPV54 (within the species Alpha-13) and HPV117 (within the species Alpha-2), respectively. MmPV4 shares <70% of L1 ORF similarity with members of the species Gamma-10 (e.g., HPV121 and HPV130) and may represent a novel species within the genus *Gammapapillomavirus*.

### Intrahost divergence of primate papillomaviruses

We focused on HPV16 because it is the most prevalent and potent carcinogen among the oncogenic HPVs [5]. To interrogate HPV16 evolution using a molecular clock, we utilized HPVs and NHP-PVs characterized in our labs and by others where the host species separation times have been well established [25, 26]. This step is essential in order to validate a vertical mutation rate model suitable for HPV variants. This model estimates the mutation rate for infectious PVs over long periods of time and might differ from horizontal mutation rates not measured in this study.

Papillomaviruses have been identified in a wide range of NHP species, including Old World monkeys and apes (e.g., macaque, chimpanzee) and New World monkeys (e.g., squirrel monkey, brown howler) [24, 27–33]. Using a maximal likelihood algorithm and a nucleotide sequence alignment of the concatenated E1-E2-L2-L1 ORFs for 141 PV types representing each species or unique host (S2 Table), we found that the majority of primate PVs phylogenetically clustered into *Alphapapillomavirus*, *Betapapillomavirus*, or *Gammapapillomavirus* genera, corresponding predominantly to the anatomical sites where the viruses were originally



**Fig 1. Phylogeny of primate papillomaviruses.** A maximum likelihood phylogenetic tree was inferred from the concatenated nucleotide sequence alignment of 4 open reading frames (E1-E2-L1-L2) of 141 papillomavirus types representing 132 species (see PV list with hosts in S2 Table). The majority of analyzed primate papillomaviruses cluster into three distinct clades, Alpha-, Beta- and Gamma-PV genera, corresponding predominately to the anatomical host sites (e.g., mucosal vs. cutaneous epithelium) where the viruses were originally isolated, rather than to the distinct host species. The branches represented by non-human primate papillomaviruses are highlighted in red. Non-primate papillomaviruses are collapsed and joined by grey lines (see comprehensive tree in S1 Fig). The dot sizes are proportional to the bootstrap percentage supports from RAxML.

<https://doi.org/10.1371/journal.ppat.1007352.g001>

isolated (e.g., mucosal or cutaneous epithelium), which was independent of the host species (Fig 1, S1 Fig and S2 Table). For example, MmPV1 is a rhesus macaque PV type (within the

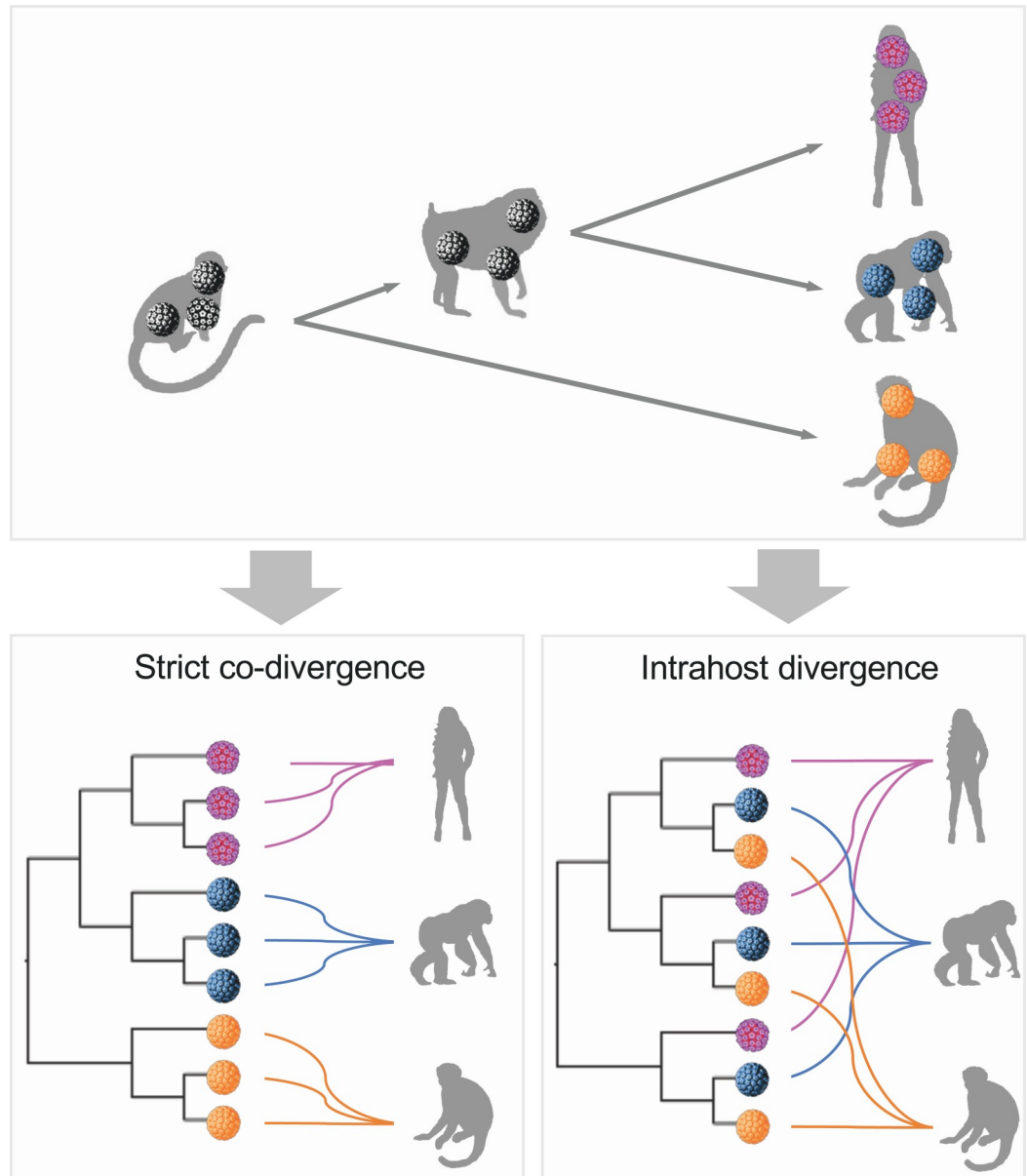
species Alpha-12) isolated from cervicovaginal cells that shares a most recent common ancestor (MRCA) with oncogenic mucosal HPV16 (within the species Alpha-9) but is distantly related to MmPV4 (within the genus *Gamma*papillomavirus), which was also isolated from a rhesus macaque. Since topological incongruence has been noted in the phylogenies of HPVs when trees are constructed with either late or early regions of the viral genomes [22, 34], we also examined the topologies of such trees. Although there was some incongruence, the majority of the primate PVs maintained their topological positions (see S2 Fig, S3 Fig and S4 Fig).

Fahrenholz's proposal for strict codivergence of host and parasites states that the "parasite phylogeny mirrors that of its host," indicating that specific pathogens isolated from an individual host species should be monophyletic to the exclusion of viruses from other host species (reviewed in de Vienne et al.) [35] (Fig 2). In the case of primate PVs, however, viruses infecting a given host species do not always cluster together, implying an ancient viral divergence model in which viral ancestors may have first split into separated viral clades corresponding to niche adaptation to specific host ecosystems (i.e., tissue tropism). Following host ancestor speciation, distinct but homophyletic viruses were transmitted to similar ecosystems (e.g., mucosal or cutaneous sites) between closely related host animals, resulting in the radiation observed in the extant primate PV tree where viruses sort by tissue tropism and not host species. This prediction was evaluated with a permutational multivariate analysis of variance (PERMANOVA) test [36] using primate PV nucleotide sequence pairwise distances, which revealed that tissue tropism (here defined by different genera) contributed to more of the variability of viral divergence (accounting for 26% of the total variance,  $p < 0.001$ ) than that of the host (6%,  $p < 0.001$ ) (Table 1).

### Co-divergence between human and non-human primate papillomaviruses

To estimate the divergence times of primate PVs from their MRCA, we used a Bayesian statistical framework employing previously established PV evolution rates [17]. Infectious PVs have been shown to have a slow mutation rate based on the observations that these double-stranded DNA viruses use the host cell DNA replication machinery, characterized by high fidelity, proofreading capacity, and post-replication repair mechanisms [37]. Since primate PVs, taken together, do not follow strict viral-host codivergence, each genus was evaluated separately to estimate divergence times. A combination of relaxed lognormal molecular clock and coalescent constant population models provided the best performance using the phylogenetic tree as shown in Fig 3A. The *Alphapapillomavirus*–*Dyoomikronpapillomavirus* split from a MRCA around 39.9 million years ago (mya) (95% highest posterior density (HPD), 36.4–43.7 mya) (Fig 3A, S2 Fig and Table 2) is consistent with the time frame of the split between New World and Old World primate ancestors [26].

Similar virus-host codivergence events were observed between Old World monkey PVs and their closest HPV relatives, and were estimated to approximately 14–31 mya (Fig 3, S5 Fig, S6 Fig and S7 Fig). For example, the species Alpha-12 (PVs mainly isolated from genital lesions of macaques) split from a MRCA with the species Alpha-9 (represented by oncogenic genital HPV16) around 27 mya coincided with the time span of the speciation between macaques and apes/humans that occurred approximately 25 mya [38, 39]. An enigmatic observation in these data is the clustering of macaque PVs (e.g., MfPV3) and baboon PV (*Papio hamadryas* PV 1, PhPV1) within the species Alpha-12 group, suggesting either a recent viral transmission between macaque and baboon monkeys, or a more complex phylogeny of the sub-family *Cercopithecinae*. The majority of distinct human PV types arose during the end of the Miocene and/or the beginning of the Pliocene epoch coincident with the divergence of humans and chimpanzees occurring around 6–8 mya (Fig 3) [40].



**Fig 2. Schematic model of virus-host codivergence.** Strict virus-host codivergence requires the evolutionary history of the pathogen to mirror that of its hosts. Clustering of viruses according to the host from which they were isolated should be observed. In addition, the divergence times of hosts and parasites should be similar (different colors highlight viruses infecting different primate host ancestors). Intrahost divergence can be defined according to specific phylogenetic criteria, such as niche-adaptation prior to coevolution in primate papillomaviruses, as opposed to clustering by hosts.

<https://doi.org/10.1371/journal.ppat.1007352.g002>

The divergence times and tree topologies support a model of intrahost divergence of primate PVs in which ancient viruses diverged and adapted to specific host ecosystems (e.g., tissue tropism or different types of epithelial cells) within an ancestral host animal lineage (e.g., the MRCA of primate animals) (Fig 4). Following periods of host speciation, continuing intrahost viral divergence events occurred as distinct but phylogenetically related viral types were transmitted to similar host ecosystems by the closely related host animals. This pattern of ancient viral divergence coupled to niche adaptation may explain, for example, the differences in the prevalence of HPV16 and HPV18 between squamous cell carcinomas and

adenocarcinomas of the cervix [41]. This difference might represent the emergence of further viral adaptation to different ecological niches within the cervix, one dominated by stratified squamous epithelium the other by columnar epithelium, respectively [42]. The fact that we do not observe similar or parallel diversity of NHP-PVs compared to HPVs (broken lines in right panel of Fig 4B) could be due, in part, to reduced sampling effort, limited population size of NHPs, bottlenecks of viral transmission, and/or restricted host migration.

### Molecular evolution and geographic distribution of HPV16 variants

Next, we constructed a phylogenetic tree of HPV16 variants based on 212 complete genomes to classify variant lineages and sublineages (S3 Table). The tree topology shows two deeply separated clades corresponding to the previously classified Eurasian and African lineages (S8 Fig), with a mean nucleotide sequence difference of  $1.72\% \pm 0.09\%$  (S4 Table). The African lineage variants were more than twice as diverse (intragroup mean difference of  $0.77\% \pm 0.04\%$ ) as the Eurasian variants ( $0.32\% \pm 0.02\%$ ). Since geographic nomenclature systems suffer from sampling biases and preconceived notions about virus ancestry, we utilized an agnostic alphanumeric nomenclature based on HPV16 phylogeny and complete genome nucleotide differences to assign HPV16 variants into four lineages designated A, B, C, and D. Each lineage could be divided into four sublineages (A1-4, B1-4, C1-4, and D1-4), based on previously described criteria (S9 Fig) [43]. The previously named Asian (As) and North American 1 (NA1) variants are designated sublineages A4 and D1, respectively [44]. The maximum pairwise difference between the most diverse isolates, from sublineages A1 and D3, was 2.23%.

Based on single-nucleotide polymorphism (SNP) patterns and phylogenetic tree topologies, we assigned 3256 HPV16 partial sequences from 22 countries/studies into variant lineages and sublineages using maximum likelihood methods (Table 3). As shown in the summarized charts of HPV16 phylogeography (Fig 5A), isolates from Asians and Caucasians (Australians/Europeans, and North Americans) were predominantly represented by A variants, with abundances of 92% and 83%, respectively. The majority of A4 variants (352/357, 99%) were from Asian individuals. Within the African population, 90% of HPV16 infections were B and C lineages. HPV16 variants in South/Central Americans were equally assigned as A1-3 (50%) and D (48%). Using a weighted UniFrac algorithm, variants were well clustered into groups (African, Eurasian, and South/Central American) corresponding to the geographic origin of the isolates (Fig 5B). Globally, A1-3 sublineages were the most widespread; whereas, the D lineages were detectable at low prevalences in many populations outside of South/Central Americans, such as in Caucasian (11%), African (7%), and Asian (6%) individuals (Fig 5C). In contrast, A4 and B/C lineages were rarely found outside of Asian and African populations, respectively.

### Divergence time estimation of HPV16 variants and other oncogenic HPV type variants

The molecular clock models used to estimate the divergence times of primate PVs support a scenario of virus-host codivergence after the virus has adapted to a specific host ecosystem.

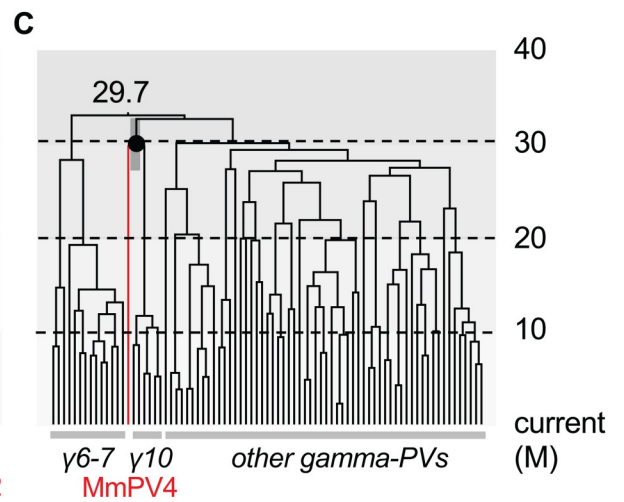
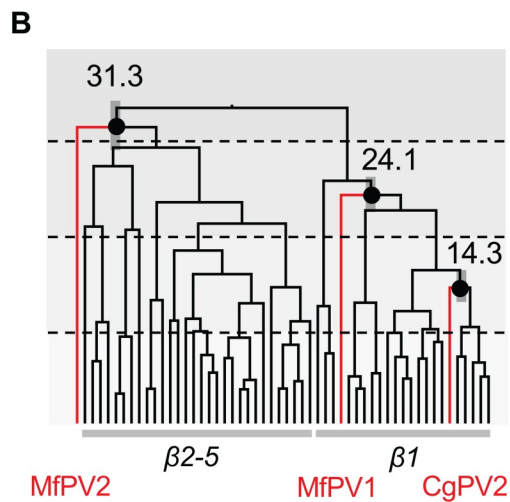
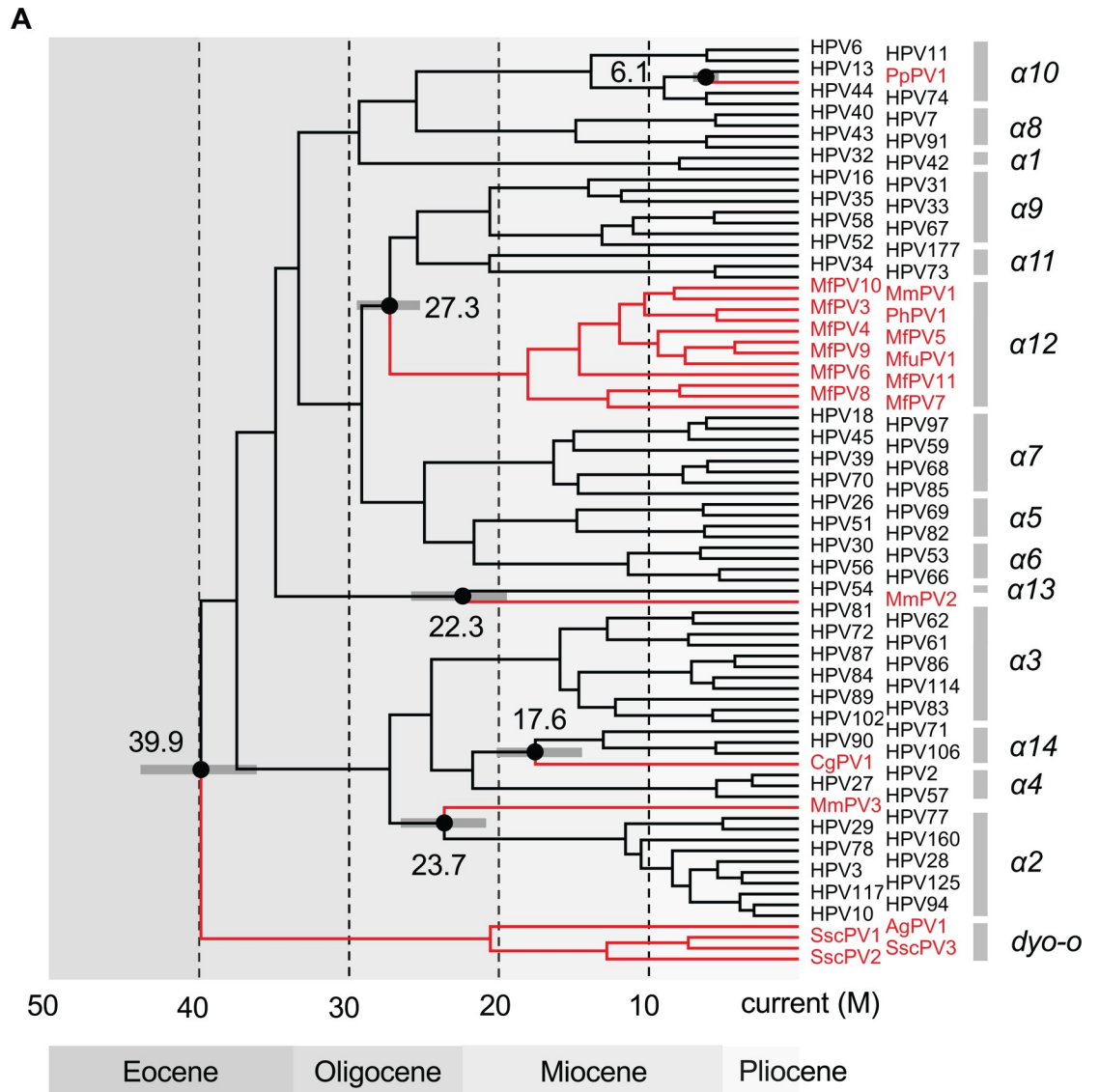
**Table 1. Permutational multivariate analysis of variance using primate papillomavirus pairwise distance.**

4ORF nt	Df	Sums of Squares	Mean of Squars	F Model	R <sup>2</sup>	Pr (>F)
Host <sup>1</sup>	1	0.3532	0.3532	4.5744	0.0606	<0.001
Genus <sup>2</sup>	4	1.5396	0.3849	4.9854	0.2641	<0.001
Residuals	51	3.9375	0.0772		0.6754	

<sup>1</sup> Host was grouped as "Human" and "Non-human primates (NHP)".

<sup>2</sup> Genus was grouped as "Alpha\_Human", "Alpha\_NHP", "Beta\_Human", "Beta\_NHP", "Gamma\_Human" and "Gamma\_NHP".

<https://doi.org/10.1371/journal.ppat.1007352.t001>





**Fig 3. Divergence time estimation of primate papillomaviruses to their most recent common ancestors (MRCAs).** A Bayesian MCMC method was used to estimate divergence times as described in the methods. Times were calculated separately for each genus, Alpha- (A), Beta- (B) and Gamma-PVs (C). Branch lengths are proportional to divergence times. The branches in red refer to non-human primate papillomaviruses. Numbers above the nodes with circles are the mean estimated divergence time in million years (M) between human and non-human papillomavirus clades. The bars in grey represent the 95% highest posterior density (HPD) interval for the divergence times (see details in S5 Fig, S6 Fig and S7 Fig, respectively). Panels B and C show time on the Y-axis and phylogeny on the X axis.

<https://doi.org/10.1371/journal.ppat.1007352.g003>

Using a similar Bayesian Markov chain Monte Carlo (MCMC) framework, we initially applied six combinations of clock models to estimate the divergence of HPV16 variants from their MRCA, without any prior assumption of virus-host codivergence (Table 4, no calibration). Interestingly, a combination of the relaxed lognormal molecular clock and coalescent Bayesian skyline models indicated that HPV16 A and BCD had divided around 618.5 thousand years ago (kya) (95% HPD: 331.5–996.1). This estimation is within the time span of the separation between *Homo sapiens* and archaic hominins (e.g., Neanderthal/Denisova) but around two–five times longer than the estimated modern *Homo sapiens* divergence time (ca. 150–200 kya) [45] indicative of an ancient divergence of HPV16 variants prior to the emergence of modern human ancestors. Based on the geographic distribution of HPV16 variants above, we then used an archaic *hominin-host-switch* (HHS) scenario to calibrate the divergence time between HPV16 A and non-A variants (500 kya, 95% HPD: 400–600), and a *modern-out-of-Africa* (MOA) scenario between BC and D variants (90 kya, 95% HPD: 60–120). When time calibrations were introduced into the phylogenetic tree, the HHS scenario showed the strongest support for time inference and estimated an initial divergence of HPV16 variants at approximately 489 kya (95% HPD: 394–581), predating the *out-of-Africa* migration of modern humans (ca. 60–120 kya) (Fig 6 and S10 Fig) [46, 47]. In addition, the demographic model of the Bayesian skyline plot for the population function through time showed a recent exponential expansion of the effective population size of present-day HPV16 occurring in the last 25 kya, lagging behind the growth of modern human populations (starting from the last 40–50 kya) (see the top panel of Fig 6). This plot most likely reflects the concurring increase and mobility of modern human populations and present-day virus populations in the last epoch.

We observed a similar divergence timeframe for other HPV variants, splitting from their MRCAs approximately 300–600 kya and showing a strong correlation between evolution times and genomic diversities (Fig 7, Table 5). In all cases, the deep separation between HPV16 variant lineages A and BCD (and the deepest lineage separations of other HPV variants) suggests an ancient virus-host codivergence, coinciding with the split between archaic Neanderthal/Denisova and modern human ancestors from their MRCA (Fig 8). Neanderthals spread out over Eurasia with at least two populations splitting approximately 77–114 kya from each other based on analysis of archaic genomes from Vindija, Mezmaiskaya (Caucasus), and Denisova (Siberia) [48]. This time period corresponds to the diversion of HPV16 A sublineages and in particular the split of A4 from A1/2/3 and the emergence of HPV16 A4 in Asia, likely representing independent transmission of A4 from archaic hominins to modern humans in the east.

## Discussion

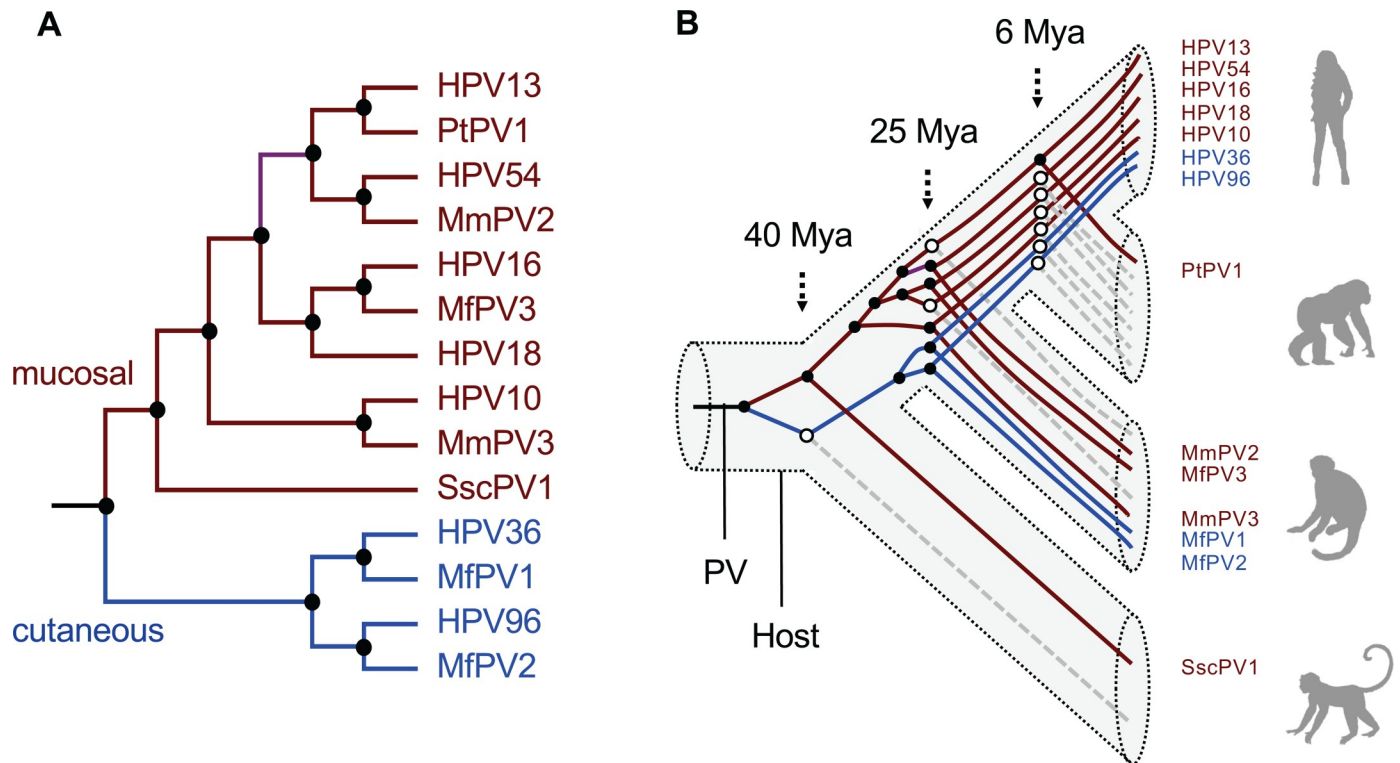
In this work, we used a Bayesian MCMC framework to estimate the divergence times of primate PVs and propose an early ancient intrahost viral divergence model (i.e., niche adaptation) followed by viral-host coevolution. This form of viral evolution has been documented for polyomaviruses [49], herpesviruses [50], and some retrovirus genera [51]. With the assumption of host niche adaptation as a fundamental process, the estimation of primate PV

Table 2. Divergence time estimation of Alphapapillomavirus and Dyoomikronpapillomavirus types.

Calibration	Clock Model	Tree Prior	AICM value	AICM difference	Log marginal likelihood	MRCA (million years ago, mya)											
						Alpha vs Dyooomega (mean)	95% HPD Interval	$\alpha 9/11$ vs $\alpha 12$ (mean)	95% HPD Interval	HPV54 vs MmPV2 (mean)	95% HPD Interval	$\alpha 2$ vs MmPV3 (mean)	95% HPD Interval	$\alpha 14$ vs CgPV1 (mean)	95% HPD Interval	HPV13 vs PpPV1 (mean)	95% HPD Interval
3 Cali.	Relaxed	Bayesian	719437	47	-359153.05	39.7	[36.2, 43.5]	27.2	[25.2, 29.3]	22.5	[19.0, 26.1]	23.8	[21.2, 26.4]	17.7	[15.7, 19.8]	6.1	[5.4, 6.8]
<b>3 Cali.#</b>	<b>Relaxed</b>	<b>Constant</b>	<b>719391</b>	-	<b>-359153.10</b>	<b>39.9</b>	<b>[36.4, 43.7]</b>	<b>27.3</b>	<b>[25.3, 29.4]</b>	<b>22.3</b>	<b>[19.1, 25.7]</b>	<b>23.7</b>	<b>[21.4, 26.1]</b>	<b>17.6</b>	<b>[15.6, 19.8]</b>	<b>6.1</b>	<b>[5.5, 6.8]</b>
3 Cali.	Relaxed	Yule	719399	8	-359153.30	39.6	[36.1, 43.2]	27.2	[25.1, 29.2]	22.2	[19.0, 25.6]	23.6	[21.3, 26.1]	17.6	[15.7, 19.7]	6.2	[5.5, 6.8]
3 Cali.	Strict	Bayesian	721072	1681	-360478.67	35.6	[32.7, 38.3]	25.9	[23.9, 27.8]	17.8	[16.3, 19.5]	21.1	[19.3, 23.0]	15.6	[14.3, 17.0]	5.3	[4.8, 5.7]
3 Cali.	Strict	Constant	721074	1684	-360478.92	35.7	[32.9, 38.5]	25.9	[24.0, 27.9]	17.9	[16.3, 19.5]	21.1	[19.4, 23.0]	15.6	[14.3, 17.0]	5.2	[4.8, 5.7]
3 Cali.	Strict	Yule	721073	1682	-360478.78	35.5	[32.9, 38.5]	25.8	[24.0, 27.8]	17.8	[16.2, 19.4]	21.1	[19.4, 22.9]	15.5	[14.3, 16.9]	5.2	[4.8, 5.7]
No	Relaxed	Bayesian	719633	242	-359275.52	33.3	[30.4, 36.1]	25.0	[23.4, 26.6]	16.9	[14.7, 19.3]	18.9	[17.5, 20.4]	13.8	[12.6, 15.1]	4.7	[4.1, 5.4]
No	Relaxed	Constant	719681	290	-359270.59	34.2	[31.6, 37.1]	24.2	[22.7, 25.7]	15.4	[13.5, 17.3]	21.0	[19.1, 22.7]	15.0	[13.5, 16.5]	4.5	[3.8, 5.2]
No	Relaxed	Yule	719636	246	-359274.64	33.2	[30.5, 36.3]	25.1	[23.5, 26.8]	17.0	[14.9, 19.2]	18.9	[17.5, 20.5]	13.9	[12.5, 15.2]	4.6	[3.9, 5.4]
No	Strict	Bayesian	721038	1647	-360471.10	31.0	[29.6, 32.3]	22.5	[21.7, 23.4]	15.6	[14.7, 16.6]	18.5	[17.6, 19.5]	13.6	[12.9, 14.3]	4.4	[4.1, 4.8]
No	Strict	Constant	721037	1646	-360471.21	31.0	[29.7, 32.4]	22.6	[21.7, 23.4]	15.6	[14.7, 16.6]	18.5	[17.5, 19.4]	13.6	[12.9, 14.3]	4.4	[4.1, 4.7]
No	Strict	Yule	721039	1648	-360471.24	31.0	[29.6, 32.3]	22.6	[21.7, 23.5]	15.6	[14.6, 16.6]	18.5	[17.5, 19.4]	13.6	[12.9, 14.3]	4.4	[4.1, 4.7]

# The best models determined by the AICM test are highlighted in bold.

<https://doi.org/10.1371/journal.ppat.1007352.t002>



**Fig 4. Schematic model of virus-host codivergence of primate papillomaviruses.** (A) A schematic topology of representative primate papillomaviruses. The branch colors represent viruses with specific host niche adaptation (brown—isolated from mucosal tissues, blue—isolated from cutaneous tissue). (B) Model of phylogeny and divergence of primate papillomaviruses. In this model, one or more primate papillomavirus ancestors evolved to colonize distinct host ecosystems prior to the speciation of a primate ancestor. A process of further viral adaptation to colonize more specific host ecosystems (represented by black circles at the nodes) may have followed upon host speciation, resulting in the radiation observed in the extant primate papillomavirus tree. The broken lines in grey (starting from open circles) represent clades for which specific HPV species lack detectable non-human primate counterparts. The 3-dimensional structure represent host phylogeny.

<https://doi.org/10.1371/journal.ppat.1007352.g004>

divergence times within niche-specific clades mirrors that of the primate host evolutionary history (Fig 4). It is clear that the evolutionary history of these well adapted, slowly evolving PVs may be significantly more complex than previously appreciated [37]. The implication of host niche adaptation of primate PVs preceding virus-host codivergence suggests a critical role for viral genetic heterogeneity and natural selection. The origin of viral genetic determinants of cervical niche adaptation further supports the hypothesis that a group of well-evolved viral genotypes also contain the determinants for cervical cancer, since this phenotype cannot exert selective pressure, as it does not support the production of infectious virus. It may also explain why a large set of cervicovaginal macaque PVs (within the species Alpha-12) associated with cervical neoplasia shares a common origin with the high-risk clade of human PVs (e.g., Alpha-9) (Fig 3A) [6, 27]. Our findings provide a framework for studying the past evolution of primate PVs infecting the genital tract niche and support a molecular clock based on phylogeny, since the generation time of PVs can only be extrapolated from empiric data based on coevolution models [17, 52].

We used this well-supported molecular clock model to estimate the divergence times of HPV16 variants. HPV16 is the most common oncogenic HPV type and shows diversity in persistence and carcinogenicity [53–55], suggesting further biological differences between variant lineages. We observed specific geographic/ethnic dispersals of HPV16 variants, such as A4 predominance in Asian populations and BCD predominance in African populations. The estimated divergence times between HPV16 A and BCD variants largely predated that of the *out-of-*

**Table 3. HPV16 variant assignment with known geographic origin.**

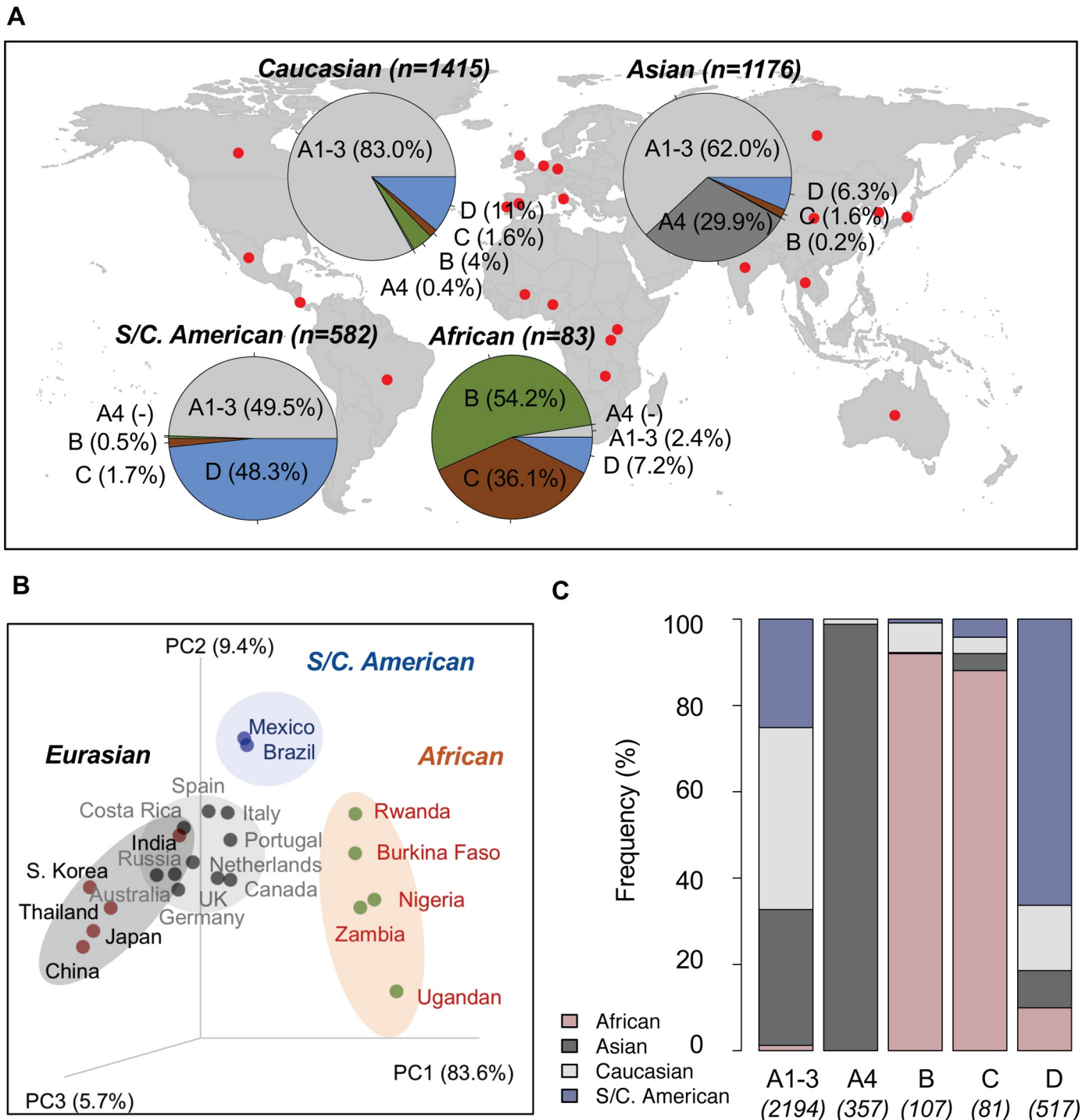
Country	Geographic/ethnic origin	Total	A1-3	A4	B	C	D	References
Burkina Faso	African	16	0	0	4	11	1	[100]
Nigeria	African	21	0	0	12	9	0	[101]
Rwanda	African	20	0	0	10	6	4	[102]
Ugandan	African	5	0	0	5	0	0	[103]
Zambia	African	21	2	0	14	4	1	[104]
China	Asian	361	148	197	2	11	3	[105–107]
India	Asian	488	427	7	0	0	54	[108, 109]
Japan	Asian	57	31	25	0	0	1	[110]
Korea	Asian	56	9	38	0	1	8	[111]
Thailand	Asian	214	114	85	0	7	8	[112, 113]
Australia	Caucasian	34	28	4	0	0	2	[132]
Canada	Caucasian	32	25	0	5	1	1	GenBank <sup>#</sup>
Costa Rica	Caucasian	693	604	1	1	2	85	[9]
Germany	Caucasian	24	24	0	0	0	0	[114]
Italy	Caucasian	197	143	0	18	4	32	[115–117]
Netherlands	Caucasian	67	59	0	3	1	4	[119, 120]
Portugal	Caucasian	187	139	0	22	8	18	[121]
Russia	Caucasian	50	48	0	0	0	2	[122]
Spain	Caucasian	75	59	0	3	1	12	[123]
United Kingdom	Caucasian	56	46	0	5	5	0	[124]
Brazil	South/Central American	200	98	0	3	10	89	[125–127]
Mexico	South/Central American	382	190	0	0	0	192	[128–131]
Total	-	3256	2194	357	107	81	517	

<sup>#</sup> GenBank/NCBI accession numbers of GQ465877 to GQ465902, and GQ479006 to GQ479011.

<https://doi.org/10.1371/journal.ppat.1007352.t003>

*Africa* migration of modern human populations, consistent with a previously reported archaic *hominin-host-switch* scenario [19, 20]. One interpretation of the data implies that the present-day Eurasian HPV16 A variants were probably the products of multiple interactions between Neanderthals/Denisovans and modern *Homo sapiens* established during sexual contact after a long period of separation (e.g., 400–600 kya). This notion of viral sexual transmission between groups is reflected in the recent genetic admixture (e.g., 80 kya) between groups [48, 56–59], with evidence of 2–4% of nuclear DNA in Eurasians that can be traced to Neanderthals [48, 58]. This assumption is likely ubiquitous in a number of Alpha-HPV variants (Fig 7, Table 5), although their pathogenesis, evolution, and epidemiology warrant further study.

Recent evidence indicates that Neanderthals spread out over the Eurasian continent and also admixed with ancestors of the present-day East Asian population [60, 61]. Since HPV16 A4 lineage is exclusively found in East Asians (approximately 40% of HPV16) and presents a higher risk of cervix cancers in Asian populations [62, 63], we speculate that a subset of Neanderthals heading east into Asia over more than 100 thousand years of existence in Eurasia could have interbred with East Asian modern humans and transmitted the HPV16 A4 sublineage and introgressed specific gene alleles that provided a selective advantage to the HPV variants coevolving with them [59, 64]. Overall, HPV16 BCD variants have higher genomic diversity than A isolates (see S4 Table), which may imply a potential population bottleneck of horizontal transmission reducing the diversity of current day A lineage isolates. In contrast, BCD variants have accumulated more genetic mutations, consistent with the observations that African populations and their pathogens have deeper origins reflected in greater diversity [65].



**Fig 5. Geographic distribution of HPV16 variants.** (A) A total of 3256 HPV16 variants with known geographic origin from 22 countries/regions (see details in Table 3) were assigned into lineage/sublineage and summarized by geographic group in the pie charts. (B) Principle component analysis using a weighted UniFrac algorithm clustered different study cohorts into three distinct groups, namely African, Eurasian (Asian and Caucasian) and South/Central American, mainly associated with a predominant population from which viruses were sampled. (C) Relative frequency of HPV16 lineages/sublineages distribute into four major geographic populations (African, Asian, Caucasian, and South/Central American).

<https://doi.org/10.1371/journal.ppat.1007352.g005>

Table 4. Divergence time estimation of HPV16 variant lineages.

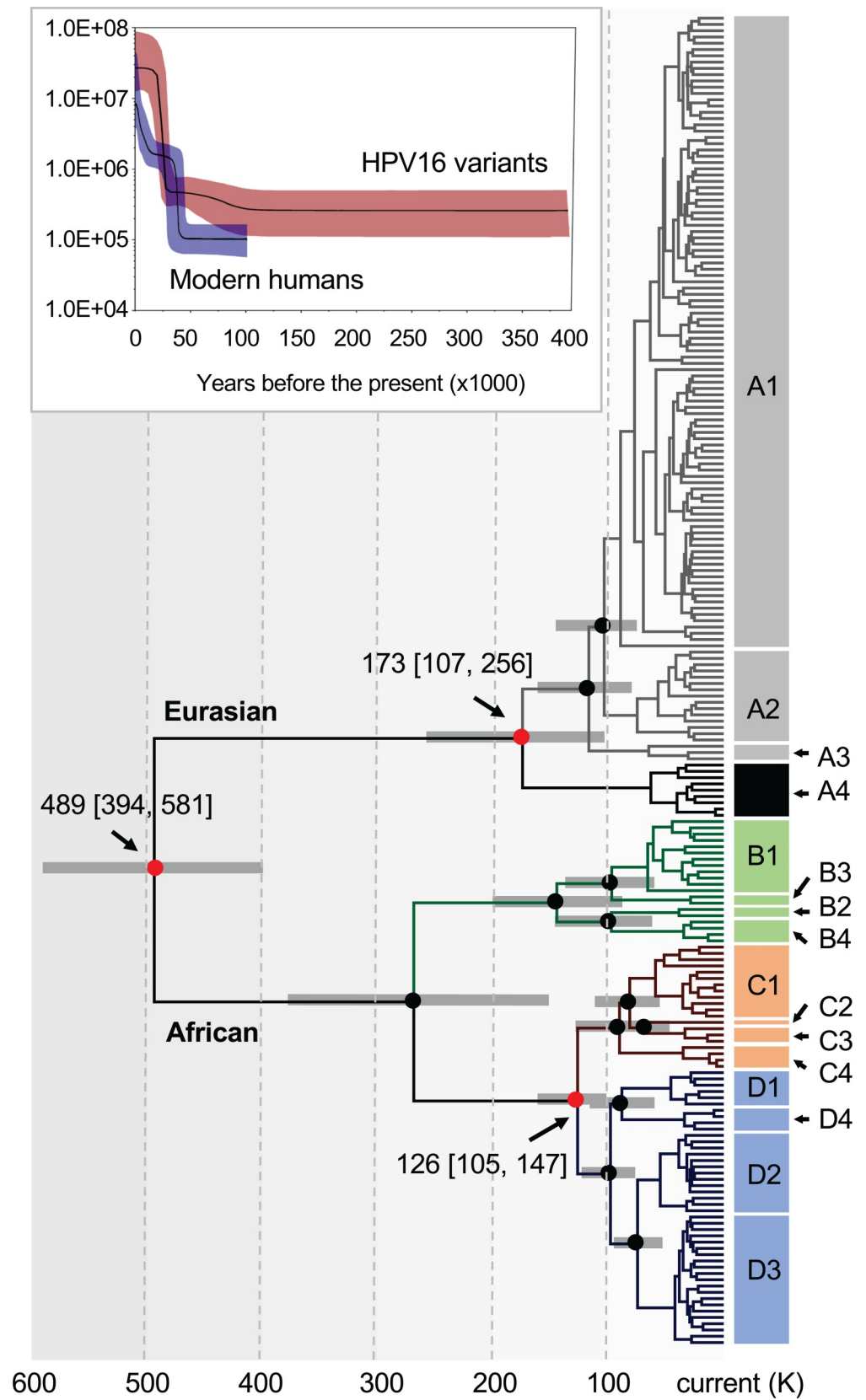
Calibration	Clock Model	Tree Prior	AICM value	AICM difference	Log marginal likelihood	MRCA (thousand years ago, kya)							
						A vs BCD (mean)	95% HPD Interval	A1/3 vs A4 (mean)	95% HPD Interval	B vs C/D (mean)	95% HPD Interval	C vs D (mean)	95% HPD Interval
No	Relaxed	Bayesian	47641	2	-23609.43	618.5	[331.5, 996.1]	199.9	[122.8, 306.4]	401.0	[250.1, 584.0]	260.0	[161.9, 381.3]
No	Relaxed	Constant	47644	4	-23606.09	1025.0	[450.4, 1843.4]	360.3	[172.0, 600.3]	667.8	[302.9, 1140.2]	439.3	[217.5, 715.0]
No	Relaxed	Yule	47655	16	-23613.10	337.2	[226.0, 460.0]	187.4	[127.1, 259.5]	286.9	[195.9, 385.3]	219.8	[141.2, 305.3]
No	Strict	Bayesian	47886	247	-23784.30	519.3	[387.8, 663.1]	177.1	[126.8, 235.1]	335.0	[246.6, 432.5]	213.1	[157.7, 277.8]
No	Strict	Constant	47852	213	-23777.28	539.8	[409.0, 686.4]	195.4	[139.4, 255.3]	353.1	[263.9, 456.3]	227.1	[168.8, 293.3]
No	Strict	Yule	47829	190	-23772.25	472.5	[355.6, 604.0]	176.4	[128.6, 230.9]	312.5	[235.2, 405.2]	204.0	[153.5, 264.3]
<b>2 Cali.<sup>#</sup></b>	<b>Relaxed</b>	<b>Bayesian</b>	<b>47639</b>	-	<b>-23609.83</b>	<b>488.9</b>	<b>[394.2, 581.0]</b>	<b>172.7</b>	<b>[106.6, 255.8]</b>	<b>266.1</b>	<b>[164.1, 386.8]</b>	<b>125.5</b>	<b>[104.8, 147.2]</b>
2 Cali.	Relaxed	Constant	47651	11	-23610.23	513.1	[421.0, 601.5]	259.7	[153.3, 383.0]	341.9	[220.2, 496.2]	132.8	[111.1, 154.2]
2 Cali.	Relaxed	Yule	47671	32	-23613.98	403.7	[316.1, 493.8]	183.0	[128.0, 253.1]	234.1	[166.8, 314.9]	130.9	[110.8, 151.5]
2 Cali.	Strict	Bayesian	47900	261	-23787.00	440.1	[384.4, 495.1]	150.9	[119.8, 184.4]	255.5	[213.7, 296.9]	142.6	[125.1, 160.7]
2 Cali.	Strict	Constant	47854	215	-23779.58	449.2	[397.0, 506.4]	165.0	[133.9, 198.3]	264.7	[222.4, 305.0]	147.6	[130.8, 165.8]
2 Cali.	Strict	Yule	47824	185	-23774.85	413.6	[362.4, 469.7]	153.3	[124.5, 184.8]	246.4	[208.3, 286.7]	143.1	[126.5, 160.4]

<sup>#</sup> The best models determined by the AICM test are highlighted in bold.

<https://doi.org/10.1371/journal.ppat.1007352.t004>

This idea supports one theory that both HPV16 BCD and modern humans arose in Africa (Fig 8). Following a relatively recent *out-of-Africa* migration, the modern humans acquired the A variant from sex with archaic hominins and possibly carried D variants into Eurasia under conditions of a small population size. The ancestors of East Asian people crossed the Bering Strait and were early populators of the Americas (based on historical records and genetic relatedness) [66]. Surprisingly, the D lineage is phylogenetically rooted in the African clade, but we did not find a major reservoir of the D lineage in the present-day African populations. This interesting observation suggests either an advantage of niche colonization and expansion of HPV16 D variants in Native Americans or a bottleneck of HPV16 variants present in people populating the Americas. Alternatively, the lack of A4 and the high proportion of D lineages in the Americans could be the result of an early colonization of the Americas by an unknown group from Africa. More data is needed to sort out the evolutionary history of the HPV16 D lineage and might provide clues to new features of the populating of the Americas.

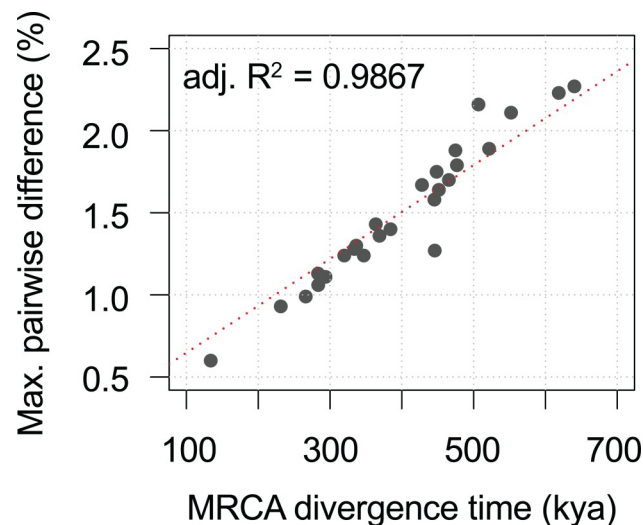
Sexual interactions between archaic hominins and modern human ancestors likely occurred over multiple time- and space-scales. For example, viral transmission might have also occurred from modern humans to Neanderthals/Denisovans, based on the evidence of ancient gene flow from early modern humans into Eastern Neanderthals [57]. Since PVs usually establish infections at the basal layer of epithelial cells, it will be impossible to detect viruses from fossil bones of archaic hominins and document the presence of HPV's in archaic hominin populations [20]. The evolutionary histories and origins of modern *H. sapiens* are undergoing



**Fig 6. Divergence time estimation of HPV16 complete genome variants.** A Bayesian MCMC method was used to calculate the divergence times of HPV16 complete genome variants from their most recent common ancestors, as described in the methods. The nodes highlighted with red circles indicate divergence times of the split between HPV16 A and non-A lineages, between A1-3 and A4 sublineages, and between C and D lineages. Branch lengths are proportional to divergence times scaled in thousands of years (K). Grey bars indicate the 95% highest posterior density (HPD) for the corresponding divergence age (see details in S10 Fig). Colors in branches represent distinct HPV16 variant lineages/sublineages. The plot on top of the tree is a Bayesian skyline estimation based on 311 present-day human mtDNA sequences (without the loop region) from geographically diverse populations and 212 HPV16 complete genome variants with a similar geographical distribution. The median posterior estimates (the product of the effective population size  $N_e$  and the generation length  $g$  in years) throughout the given time period are illustrated with lines in black. The dark blue (humans) and dark red (HPV16) areas give the 95% HPD interval of these estimates.

<https://doi.org/10.1371/journal.ppat.1007352.g006>

dramatic revisions with the introduction of advanced sequencing techniques and methods to analyze genomic samples from archaic hominin specimens [67–69]. Since the reproductive success per copulation between *H. sapiens* and archaic hominins is predicted to have lower viability than that of modern human reproductive events, high levels of sexual interaction were likely present facilitating HPV transmission, in addition to genetic introgression observed in modern non-African populations [70]. For example, the human leukocyte antigen (HLA) B\*07:02 and C\*07:02 alleles associated with increased risk in cervix cancers appear to be introgressed regions in present-day Eurasians and Melanesians from Neanderthals or Denisovans [71–73]. This also suggests that adaptive introgression of modern humans from archaic hominins influences the pathogenic outcome of these infections by as yet unknown mechanisms [70, 74]. However, it can be speculated that introgressed genes providing some selective advantage to hybrid human-archaic hominin offsprings could also make them more susceptible to HPV variants adapted to archaic hominins over hundreds of thousands of years of coevolution. The introgressed genes are most likely related to immunity against infections, whatever the pathogens might be and HPV was along for the ride, since HPV is not known to affect reproductive fitness of the host.



**Fig 7. Ancient HPV variant codivergence with archaic hominins.** The plot shows the correlation between divergence time (X-axis) of HPV variants from the most recent common ancestors and genomic diversity (Y-axis) of HPV variants. The Alpha-3 (HPV61), Alpha-5 (HPV26, 51, 69, 82), Alpha-6 (HPV30, 53, 56, 66), Alpha-7 (HPV18, 39, 45, 59, 68, 70, 85, 97), Alpha-9 (HPV16, 31, 33, 35, 52, 58, 67), Alpha-10 (HPV6, 11), Alpha-11 (HPV34, 73), and Alpha-13 (HPV54) variants from previous publications were included. The adjusted  $R^2$  value indicating the correlation between sequence diversity and divergence time of HPV type variants was calculated using the linear model (*lm*) function in R.

<https://doi.org/10.1371/journal.ppat.1007352.g007>



Table 5. Estimation of divergence time (thousand years ago, kya) of HPV variants from the most recent common ancestor (MRCA).

HPV type	Number of isolates	MRCA divergene (kya)		Estimated Rate (10x <sup>-8</sup> )		Maximum nuc. difference		Variant lineage/ sublineage	Deepest separated clades
		Mean	95% HPD Interval	Mean	95% HPD Interval	Mean	s.d.		
HPV61	10	640.2	[280.9, 1101.7]	1.77	[0.94, 2.60]	2.27%	0.17%	A1-2, B, C	A/B vs C
HPV16	212	618.5	[331.5, 996.1]	2.09	[1.85, 2.31]	2.23%	0.16%	A1-4, B1-4, C1-4, D1-4	A vs B/C/D
HPV18	46	552.1	[428.5, 704.6]	1.80	[1.61, 1.96]	2.11%	0.16%	A1-5, B1-3, C	A vs B/C
HPV30	16	521.5	[340.4, 717.7]	1.79	[1.43, 2.12]	1.89%	0.17%	A1-5, B	A vs B
HPV52	150	506.7	[203.7, 989.6]	1.84	[1.41, 2.26]	2.16%	0.17%	A1-2, B1-3, C1-2, D	A/B/C vs D
HPV53	27	476.7	[404.8, 549.7]	1.84	[1.73, 1.92]	1.79%	0.13%	A, B, C, D1-4	A/B vs C/D
HPV68CDEF	15	474.6	[356.6, 614.4]	1.86	[1.38, 2.34]	1.88%	0.12%	C1-2, D1-2, E, F1-2	C/F vs D/E
HPV58	138	465.4	[315.4, 659.8]	1.74	[1.53, 1.95]	1.70%	0.17%	A1-3, B1-2, C, D1-2	A vs B/C/D
HPV70	9	451.5	[336.7, 572.7]	1.82	[1.51, 2.09]	1.64%	0.12%	A, B	A vs B
HPV66	14	448.4	[257.7, 638.7]	1.83	[1.36, 2.30]	1.75%	0.14%	A, B1-2	A vs B
HPV54BC	5	445.8	[162.9, 845.7]	1.88	[0.84, 3.12]	1.27%	0.13%	B, C1-2	B vs C
HPV6	190	445.3	[366.0, 528.4]	1.84	[1.80, 1.88]	1.58%	0.13%	A, B1-5	A vs B
HPV45	33	428.0	[288.3, 577.8]	1.77	[1.47, 2.07]	1.67%	0.15%	A1-3, B1-3	A vs B
HPV11	78	384.2	[279.2, 483.3]	1.85	[1.67, 2.02]	1.40%	0.15%	A1-4, B	A vs B
HPV31	88	368.8	[230.9, 541.3]	1.76	[1.55, 1.96]	1.36%	0.14%	A1-2, B1-2, C1-4	A/B vs C
HPV73	16	363.6	[175.9, 596.6]	1.77	[1.08, 2.41]	1.43%	0.12%	A1-2, B	A vs B
HPV34AB	8	346.9	[270.2, 421.5]	1.84	[1.65, 2.03]	1.24%	0.11%	A1-2, B	A vs B
HPV59	8	336.5	[259.8, 413.6]	1.83	[1.61, 2.05]	1.30%	0.13%	A1-3, B	A vs B
HPV51	29	333.8	[242.9, 429.1]	1.82	[1.56, 2.04]	1.28%	0.12%	A1-4, B1-2	A vs B
HPV33	22	319.7	[211.3, 441.4]	1.81	[1.44, 2.13]	1.24%	0.12%	A1-2, B, C	A vs B/C
HPV39	20	293.6	[219.3, 374.2]	1.83	[1.61, 2.02]	1.11%	0.11%	A1-2, B	A vs B
HPV67	8	283.5	[219.2, 350.1]	1.83	[1.64, 2.00]	1.06%	0.11%	A1-2, B	A vs B
HPV82AB	11	283.3	[202.2, 369.3]	1.89	[1.37, 2.47]	1.13%	0.10%	A1-3, B1-2	A vs B
HPV56	12	266.1	[177.5, 354.9]	1.80	[1.44, 2.13]	0.99%	0.11%	A1-2, B	A vs B
HPV69	7	231.3	[173.9, 294.7]	1.81	[1.45, 2.07]	0.93%	0.10%	A1-4	A1/2 vs A3/4
HPV35	24	133.8	[98.0, 173.9]	1.81	[1.61, 1.99]	0.60%	0.07%	A1-2	A1 vs A2

Time was estimated using uncorrelated lognormal distribution molecular clock and coalescent Bayesian skyline models without calibration.

<https://doi.org/10.1371/journal.ppat.1007352.t005>

This study has its strengths and limitations. We expand the current understanding of HPV16 evolution beyond the recent description of HPV transmission between archaic and modern humans that used existing data [20] in important ways. We have expanded the understanding of HPV16 in the context of human and non-human primate PV evolution by characterizing additional New World and Old World monkey PVs and using the known divergence times of specific primate species to establish a valid molecular clock. This approach was used to establish the times of Neanderthal divergences [48]. We demonstrate that niche adaptation had to proceed viral-host coevolution, and suggest that subsequent niche adaptation might underlie the difference in prevalence of HPV16 and HPV18 in cervical squamous and glandular lesions. We have identified and characterized additional HPV16 variants enabling us to establish the HPV16 variant taxonomy that includes subvariants that have unique biological characteristics [53]. Moreover, we propose that evolution of HPV16 A in Neanderthals over time led to allopatric emergence of the HPV16 A4 lineage as Neanderthals moved east and interbred with modern humans in Asia. We have also expanded the number of HPV16 isolates from around the world to establish the global distribution of HPV16 variants. Lastly, we provide new interpretations and questions on the HPV16 D lineage that is part of the African

clade, but is highly prevalent in South/Central America. Nevertheless, there are also limitations to the current study and interpretations. The understanding of human evolution is constantly being challenged with new data and it is possible the models of human evolution used in this study will change [75]. We have not sampled every population and it is possible that additional HPV16 isolate data could change our interpretations. The data obtained on the geographic locations of the HPV partial sequences could be incorrect resulting in underestimating the true associations between variants and historic origins. Lastly, it is possible that very low population sizes of humans migrating out of Africa carried HPV16 A lineage variants leaving no traces in Africa, but expanding throughout Eurasia. This unlikely possibility would influence the interpretations of both our work and that of previous studies analyzing the evolution of HPV16 [20].

In conclusion, the biology and natural life cycle of oncogenic HPVs that results in infectious viral particles (i.e., vegetative virus life cycle) is highly adapted to the differentiation program of epithelial cells [76]. Poorly differentiated precancerous and cancerous cells in the cervix do not support the HPV vegetative life cycle, and thus viral-associated transformation does not contribute to the fitness of HPVs. Viral phenotypes that serve to adapt to a specific ecological niche, evade host immune mechanisms, and support persistent viral production, however, should contribute to viral fitness. Therefore, further investigations of viral-host interactions and the underlying mechanisms of viral oncogenicity, should continue to focus on features of viral evolution and niche adaptation that contribute to fitness, since the oncogenic outcome of HPV infections appear to be “collateral damage” affecting host morbidity and mortality. The current data provides a framework to unravel the mysteries of oncogenic HPV genomes as we expand our understanding of viral-host evolution.

## Materials and methods

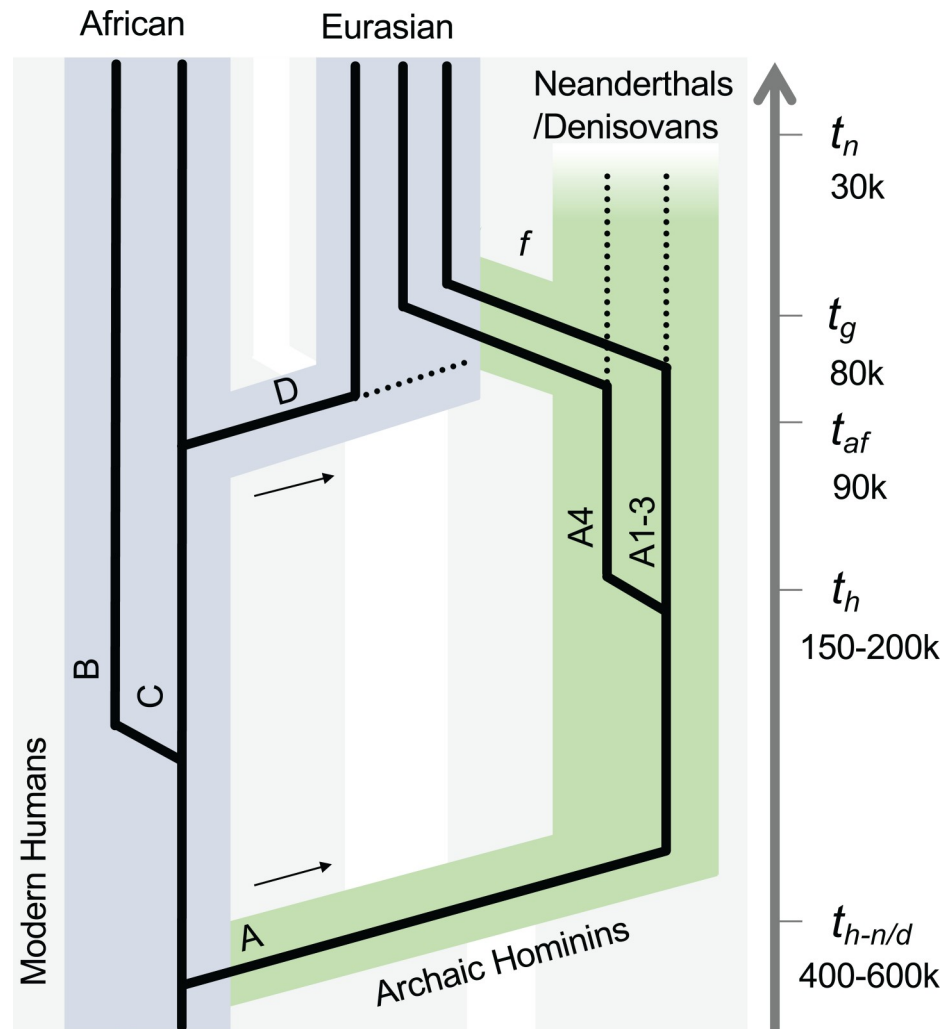
### Ethics statement

The studies providing human cellular samples have been approved by the Institution Review Board of the Albert Einstein College of Medicine, Bronx, NY, and the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee. All human subjects were older than 18 years of age and samples were anonymized without individual identifying information. Written informed consent was obtained from each participant.

The animal use protocol was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of Albert Einstein College of Medicine (protocol number 20060908). All procedures involving animals were conducted in compliance with applicable state and federal laws, guidelines established by the Animal Care and Use Committees of the respective institutions, and standards of the U.S. Department of Health and Human Services, including the National Institutes of Health Guide for the Care and Use of Laboratory Animals. The programs for animal care and welfare at Albert Einstein College of Medicine has been fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). The Animal Welfare Assurance (A3312-01) is on file with the Office for Laboratory Animal Welfare.

### *Saimiri sciureus* and *Macaca mulatta* papillomavirus isolates and complete genome characterization

The *Saimiri sciureus* PV DNA was isolated from exfoliated cervical cells of two adult female squirrel monkeys screened using polymerase chain reaction (PCR)-based MY09/11 and FAP59/64 primer systems [77, 78]. Sequences from the PCR products were compared with a



**Fig 8. Schematic illustration of HPV16 codivergence with archaic hominins.** The model was based on HPV16 variant divergence time estimation, phylogenetic topology, and geographic distribution that superimposes an ancestral viral transmission between Neanderthals/Denisovans and modern human populations. The early divergence event among deeply separated HPV16 variant lineages (A vs. BCD) suggests ancient virus-host codivergence following the speciation of modern humans and archaic hominins (e.g., Neanderthals and Denisovans) from their most recent common ancestors. The gene flow through host interbreeding between archaic hominins allowed viral transmission from Neanderthals/Denisovans to modern humans.  $t_{h-n/d}$  denotes the splitting time between Neanderthals/Denisovans and modern humans,  $t_h$  represents the speciation of modern humans.  $t_{af}$  indicates the era of population expansion of modern humans walking out-of-Africa.  $t_g$  indicates the time of gene flow (f) that may have occurred between modern humans and Neanderthals/Denisovans.  $t_n$  estimates the extinction of Neanderthals/Denisovans. The arrows indicate the out-of-Africa migration events of archaic and modern human populations. The broken lines indicate potential extinction of viral variants. Branch lengths and widths are not drawn to scale.

<https://doi.org/10.1371/journal.ppat.1007352.g008>

PV database maintained in the Burk lab using a Blastn search and shown to have < 90% similarities to previously characterized PV types. The whole genomes were PCR-amplified as two overlapping fragments using degenerate primer sets designed on available L1 gene sequences and consensus E1 alignments, and subsequently Sanger sequenced using primer walking in the Einstein Sequencing Facility, New York [33]. Geneious R9.1.7 was used to assemble segmented sequences into the complete genome sequences and identify ORFs [79].

The *Macaca mulatta* PV DNA was purified from exfoliated cervical cells of one adult female rhesus monkey and swabs from the penis surface of one adult male rhesus monkey. The viral

DNA was initially detected using multiplexed next-generation sequencing (NGS) assays targeting two small fragments (136 bp and 83 bp, respectively) within the L1 ORF [80, 81]. Sequences of a Blastn search against a PV database showed < 90% similarities to characterized PV types. The total DNA underwent a metagenomic sequencing on an Illumina HiSeq4000 at Weill Cornell Medicine Genomics Resources Core Facility, New York, using paired-end 100 bp reads. The short reads were filtered for host genome contamination and assembled *de novo* using Megahit v1.0.6 to build contigs with long length [82]. The whole genomes of novel *Macaca mulatta* PVs were validated using type-specific PCR in three overlapping fragments and Sanger sequencing using a primer walking strategy.

The complete genome sequences of SscPV1/2/3 and MmPV2/3/4 have been submitted to NCBI/GenBank database, with access numbers of JF304765 to JF304767 and MG837557 to MG837559, respectively.

### Human papillomavirus type 16 complete genome sequencing

In our previous work, we sequenced the complete genomes of 78 HPV16 isolates (see HPV16 list in S3 Table) [83, 84]. In the current study, 122 cervicovaginal samples containing HPV16 DNA were randomly chosen from the Kaiser Permanente Northern California (KPNC)-NCI HPV Persistence and Progression (PaP) cohort study [85] and a population-based HPV prevalence survey coordinated by the International Agency for Research on Cancer (IARC) [63]. The complete genomes were characterized using nested overlapping PCR and Sanger sequencing as previously reported [86]. The PaP study samples were also sequenced using Ion PGM platform [87]. In addition, 12 HPV16 complete genomes sequenced by others were included in this study [88–92].

### Phylogenetic analyses and tree construction

To evaluate the phylogenetic relationships of PVs, the concatenated nucleotide sequences of four open reading frames (ORFs) of the E1, E2, L2, and L1 genes of 141 PV types representing 132 species and unique hosts were used (see PV list in S2 Table, column labelled “Selected type” marked yes). Because all known PVs contain these four core ORFs, the concatenated partitions provide a comprehensive evaluation of the evolutionary history of *Papillomaviridae*. In addition, the highly conserved E1 early gene and L1 late gene were used to characterize phylogenetic incongruence. The nucleotide sequences of each coding region were aligned based on the corresponding amino acid sequences previously aligned using MUSCLE v3.8.31 [93] in Geneious R9.1.7. For HPV16 lineage/sublineage classification and phylogenetic analyses, all 212 complete genome nucleotide sequences (see HPV16 list in S3 Table) were linearized at the ATG of the E1 ORF and aligned using MAFFT v7.221 [94].

Maximum likelihood (ML) trees were constructed using RAxML MPI v8.2.3 [95] and PhyML MPI v3.1 [96] with optimized parameters based on the aligned complete genome nucleotide sequences. Data were bootstrap resampled 1,000 times in RAxML and PhyML. MrBayes v3.1.2 [97] with 10,000,000 cycles for the Markov chain Monte Carlo (MCMC) algorithm was used to generate Bayesian trees. A 10% discarded burn-in was set to eliminate iterations at the beginning of the MCMC run. The average standard deviation of split frequencies was checked to confirm the independent analyses approach stationarity when the convergence diagnostic approached <0.001 as runs converge. For Bayesian tree construction, the computer program ModelTest v3.7 [98] was used to identify the best evolutionary model; the identified General Time Reversible (GTR) model was set for among-site rate variation and allowed substitution rates of aligned sequences to be different. The CIPRES Science Gateway [99] was accessed to facilitate RAxML and MrBayes high-performance computation.

Permutational multivariate analysis of variance was performed using the *adonis* function in R's package 'vegan' and the pairwise distance based on 220 primate papillomavirus E1-E2-L2-L1 nucleotide sequences (S2 Table).

### Geographic dispersal of HPV16 variants worldwide

A dataset of 3256 partial sequences spanning variable genes/regions of HPV16 was obtained from GenBank that included the geographic source of the sequences mainly from indigenous ethnicities and/or local communities including 22 countries/regions throughout the world. These included, in Africa: Burkina Faso [100], Nigeria [101], Rwanda [102], Uganda [103], and Zambia [104]; in Asia: China [105–107], India [108, 109], Japan [110], Korea [111], and Thailand [112, 113]; in Europe: Germany [114], Italy [115–118], Netherland [119, 120], Portugal [121], Russian [122], Spain [123], and United Kingdom [124]; in North America: Canada (GenBank, see details in Table 3), Costa Rica [9]; in South/Central America: Bazile [125–127] and Mexico [128–131]; and Australia [132] (see Table 3). We used a maximum phylogenetic likelihood algorithm in pplacer v1.1.alpha17 [133] to place partial sequences on a reference tree inferred from an alignment composed of the 212 HPV16 variant complete genomes described in this study. A cutoff value of maximum likelihood  $\geq 0.8$  was set as confident assignment of HPV16 isolates into lineages and sublineages. The abundance of each lineage from the same country was combined and normalized using a percentage. According to the geographic patterns of HPV16 variants [44], four ethnical groups, namely African, Asian, Caucasian, and South/Central American, were summarized; for each HPV16 (sub)lineage, its frequency in each group was calculated based on the summary of individual percent abundance divided by the summary of total percent abundance. We used a weighted UniFrac method in R's package 'GUniFrac' [134] to calculate the pairwise distances between geographic locations, based on which a principle component analysis (PCoA) was performed to visualize the clustering of geographic groups of HPV16 variants using the *betadisper* function in R's package 'vegan'.

### Estimation of divergence times

We used a Bayesian Markov Chain Monte Carlo (MCMC) method implemented by BEAST v2.4.5 [135] and the previously published PV evolutionary rates [17] to estimate the divergence times of primate PVs from their most recent common ancestors (MRCAs). Times were calculated separately for *Alphapapillomavirus* ( $n = 85$ ), *Betapapillomavirus* ( $n = 54$ ), and *Gammapapillomavirus* ( $n = 81$ ) (S2 Table), given that primate PVs, taken together, do not follow strict virus-host codivergence. Three tree priors were estimated using the following demographic models: (1) coalescent constant population, (2) Yule model, and (3) coalescent Bayesian skyline, with assumptions that (1) the PV genome has a strict mutation rate or (2) there is an uncorrelated lognormal distribution (UCLD) molecular clock model of rate variation among branches, resulting in six combinations of models. In addition, we chose the GTR sequence evolution model with the gamma-distributed rate heterogeneity among sites and a proportion of invariant sites (GTR + G + I) determined by the best-fit model approach of Modeltest v3.7 [98]. The concatenated nucleotide sequence partitions of six ORFs (E6, E7, E1, E2, L2, and L1) with variable rates of substitution over time were used:  $2.39 \times 10^{-8}$  (95% confidence interval  $1.70\text{--}3.26 \times 10^{-8}$ ) substitutions per site per year for the E6 gene,  $1.44 \times 10^{-8}$  ( $0.97\text{--}2.00 \times 10^{-8}$ ) for the E7 gene,  $1.76 \times 10^{-8}$  (95% CI:  $1.20\text{--}2.31 \times 10^{-8}$ ) for the E1 gene,  $2.11 \times 10^{-8}$  (95% CI:  $1.52\text{--}2.81 \times 10^{-8}$ ) for the E2 gene,  $2.13 \times 10^{-8}$  (95% CI:  $1.46\text{--}2.76 \times 10^{-8}$ ) for the L2 gene, and  $1.84 \times 10^{-8}$  (95% CI:  $1.27\text{--}2.35 \times 10^{-8}$ ) for the L1 gene, as previously described [17]. In order to calibrate the divergence times, we introduced three time points inside and at the root of the

*Alphapapillomavirus* tree, with assumptions of codivergence histories between primate PVs and their hosts: (1) the node between HPV13 and chimpanzee PpPV1 (*Pan paniscus* PV 1) at 7 mya (95% CI, 6–8 mya) matching the split between hominin and chimpanzee ancestors; (2) the node between the species Alpha-12 (represented by *Macaca mulatta* PV 1) and Alpha-9/11 (represented by HPV16) at 28 mya (25–31 mya) matching the speciation between hominin and macaque ancestors; and (3) the node between *Alphapapillomavirus* and *Dyoomikronpapillomavirus* (represented by *Saimiri sciureus* PV 1) at 49 mya (41–58 mya) matching the divergence between Old World and New World monkey ancestors [26]. For *Betapapillomavirus* and *Gammapapillomavirus* trees, the calibration time point(s) was set between macaque PVs and their closet HPV relatives.

To estimate divergence times of HPV16 complete genome variants, a *Hominin-host-switch* (HHS) model assuming there was an ancestral viral transmission between archaic and modern human populations [20] was applied by setting two evolutionary time points to calibrate the HPV16 variant phylogenetic tree: (1) the archaic divergence of modern humans and Neanderthals/Denisovans around 500 thousand years ago (kya) (95% CI, 400–600 kya) [136] matching the split between HPV16 Eurasian (A) and African variants (BCD), and (2) the modern human *out-of-Africa* migration at 90 kya (95% CI, 60–120 kya) [45, 137], locating the era when HPV16 D variants diverged from their most recent common ancestor (MRCA). A HPV16 variant substitution rate was used for validation of a uniform prior rate:  $1.84 \times 10^{-8}$  (95% CI,  $1.43\text{--}2.21 \times 10^{-8}$ ) [20], with combinations of three tree priors and two clock models as described above. Due to the lack of geographic/ethnic dispersal information of other HPV type variants, we estimated the youngest divergence events splitting from their MRCA using complete genome alignments and HPV16 variant substitution rate without time point calibration.

To compare the population dynamics of HPV16 variants and the modern human host, Bayesian skyline plots were created using BEAST. A total of 311 globally sampled present-day human mitochondrial DNA (mtDNA) sequences, excluding the 1120 bp non-coding D-loop (that evolves at a higher rate) to give an alignment of 15,471 bp in length [138], were analyzed using a strict clock model and a coalescent Bayesian skyline, with an estimated rate of  $2.47 \times 10^{-8}$  (95% CI,  $2.16\text{--}3.16 \times 10^{-8}$ ) substitutions per site per year [139], as these sequences have been shown to evolve in a roughly clock-like manner [140, 141]. Two evolutionary time points were used to calibrate the modern human mtDNA tree: (1) the age of the MRCA between the maximum distanced modern humans, estimated to be  $171,500 \pm 50,000$  years ago, and (2) the age of the MRCA of the youngest clade that contains both African and non-African lineages, approximately  $52,000 \pm 27,500$  years ago [140].

The MCMC analysis was run for 100,000,000 steps, with subsampling every 10,000 generations. A discarded burn-in of the first 10% steps was set to refine trees and log-files for further analysis. Effective sample sizes (ESS) of all parameters are  $>300$  (*Alphapapillomavirus* tree) and  $>2000$  (HPV variant trees of each type), indicating that all Bayesian chains were well sampled and have converged. Best model estimates were selected using a posterior simulation-based analogue of Akaike's Information Criterion for MCMC samples (AICM) [142], as implemented in Tracer v.1.6. The lower AICM values indicated a better model fit. A consensus tree was inferred using TreeAnnotator v.2.4.5 and visualized using scripts developed in-house in R. The linear model (*lm*) function in R was used to estimate the correlation between sequence diversity and divergence time of HPV types and variants.

## Supporting information

**S1 Fig. Phylogeny of papillomaviruses inferred from concatenated E1, E2, L2 and L1 genes.** A maximum likelihood phylogenetic tree inferred from the concatenated nucleotide

sequence alignment of 4 open reading frames (E1-E2-L1-L2) of 141 papillomavirus types representing 132 species (see PV list in [S2 Table](#), column of “Selected type”). The main clades containing the majority of primate papillomavirus species are highlighted in grey.

(TIF)

**S2 Fig. Phylogenetic incongruence of papillomaviruses.** Maximum likelihood phylogenetic trees were inferred from the nucleotide sequence alignment of E1 (left) and L1 ORFs (right) of 141 papillomavirus types representing 132 species (see PV list with hosts in [S2 Table](#)). Although phylogenetic incongruence was observed between trees based on individual genes, the classification of the majority of characterized primate PVs largely corresponds to the grouping based on tissue tropism and biological characteristics. The branches represented by non-human primate papillomaviruses are highlighted in red. Non-primate papillomaviruses are collapsed and joined by grey lines (see comprehensive tree in [S3 Fig](#) and [S4 Fig](#)). The dot sizes are proportional to the bootstrap percentage supports from RAxML.

(TIF)

**S3 Fig. Phylogeny of papillomaviruses inferred from E1 gene.** A maximum likelihood phylogenetic tree inferred from the nucleotide sequence alignment of E1 gene of 141 papillomavirus types representing 132 species (see PV list in [S2 Table](#), column of “Selected type”). The main clades containing the majority of primate papillomavirus species are highlighted in grey.

(TIF)

**S4 Fig. Phylogeny of papillomaviruses inferred from L1 gene.** A maximum likelihood phylogenetic tree inferred from the nucleotide sequence alignment of L1 gene of 141 papillomavirus types representing 132 species (see PV list in [S2 Table](#), column of “Selected types”). The main clades containing the majority of primate papillomavirus species are highlighted in grey.

(TIF)

**S5 Fig. Divergence time estimation of Alphapapillomaviruses and Dyoomikronpapillomavirus to their most recent common ancestors (MRCAs).** A Bayesian MCMC method was used to estimate divergence times as described in the methods. Branch lengths are proportional to divergence times. The branches in red refer to non-human primate papillomaviruses. Numbers above the nodes with circles are the mean estimated divergence times in millions of years (M) between human and non-human papillomavirus clades. The bars in grey represent the 95% highest posterior density (HPD) interval for the divergence times. The viral genomes included can be found in [S2 Table](#).

(TIF)

**S6 Fig. Divergence time estimation of Betapapillomaviruses to their most recent common ancestors (MRCAs).** A Bayesian MCMC method was used to estimate divergence times as described in the methods. Branch lengths are proportional to divergence times. The branches in red refer to non-human primate papillomaviruses. Numbers above the nodes with circles are the mean estimated divergence times in millions of years (M) between human and non-human papillomavirus clades. The bars in grey represent the 95% highest posterior density (HPD) interval for the divergence times. The viral genomes included can be found in [S2 Table](#).

(TIF)

**S7 Fig. Divergence time estimation of Gammapapillomaviruses to their most recent common ancestors (MRCAs).** A Bayesian MCMC method was used to estimate divergence times as described in the methods. Branch lengths are proportional to divergence times. The branches in red refer to non-human primate papillomaviruses. Numbers above the nodes with

circles are the mean estimated divergence times in millions of years (M) between human and non-human papillomavirus clades. The bars in grey represent the 95% highest posterior density (HPD) interval for the divergence times. The viral genomes included can be found in [S2 Table](#).

(TIF)

**S8 Fig. HPV16 complete genome tree topology.** Maximum likelihood trees of HPV16 variant isolates inferred from 212 complete genomes listed in [S3 Table](#). Variant lineages (e.g., termed A and B, etc.) and sublineages (e.g., termed A1 and A2, etc.) are named using an alphanumeric nomenclature system. Inter-sublineage bootstrap supports by PhyML and RAxML are labeled at the key nodes. Colors represent different HPV16 lineages. The bar indicates the nucleotide substitution of unit changes per site.

(TIF)

**S9 Fig. Heatmap of HPV16 complete genome pairwise diversity.** Pairwise sequence identity based on the nucleotide sequence alignment of 212 HPV16 complete genomes was measured and represented as a heatmap and scaled such that the maximum inter-sequence identity differences (2.23%) are displayed as red and the minimum inter-sequence identity differences (0.00%) as blue.

(TIF)

**S10 Fig. Divergence time estimation of HPV16 complete genome variants.** A Bayesian MCMC method was used to calculate the divergence times of HPV16 complete genome variants from their most recent common ancestors as described in the methods. A previously published HPV16 variant substitution rate and two human evolutionary time points of calibration (red circles) were set. Branch lengths are proportional to the times and are scaled in millions of years (M). Grey bars indicate the 95% highest posterior density (HPD) for the corresponding divergence age. Colors in branches represent distinct HPV16 variant lineages.

(TIF)

**S1 Table. Novel papillomaviruses isolated from squirrel monkeys (*Saimiri sciureus*) and rhesus monkeys (*Macaca mulatta*).**

(XLSX)

**S2 Table. List of papillomavirus types used in this study.**

(XLSX)

**S3 Table. List of 212 HPV16 complete genome variants.**

(XLSX)

**S4 Table. Nucleotide sequence mean difference ( $\pm$  standard error) of HPV16 complete genome variants between and within lineages/sublineages.**

(XLSX)

## Acknowledgments

We would like to acknowledge the large number of laboratory members that contributed to multiple studies included in this work. We thank Drs. Koenraad van Doorslaer and Benjamin Smith for advice when preparing the original manuscript. We thank Dr. Luciana Bueno de Freitas and members of the Burk laboratory for performing HPV DNA genotyping analyses and Sanger sequencing. We also thank Miss Po Yee Wong, Miss Wendy Ho, Mr. Samuel Tong, and Mr. Jeffrey Xu for the assistance with rhesus monkey PV sequencing.



## Author Contributions

**Conceptualization:** Zigui Chen, Robert D. Burk.

**Data curation:** Zigui Chen, Rob DeSalle, Gary M. Clifford, Robert D. Burk.

**Formal analysis:** Zigui Chen.

**Funding acquisition:** Zigui Chen, Gary M. Clifford, Robert D. Burk.

**Investigation:** Zigui Chen, Robert D. Burk.

**Methodology:** Zigui Chen, Rob DeSalle, Robert D. Burk.

**Project administration:** Zigui Chen, Robert D. Burk.

**Resources:** Zigui Chen, Mark Schiffman, Rolando Herrero, Charles E. Wood, Julio C. Ruiz, Gary M. Clifford, Paul K. S. Chan, Robert D. Burk.

**Software:** Zigui Chen.

**Supervision:** Zigui Chen, Robert D. Burk.

**Validation:** Zigui Chen, Robert D. Burk.

**Visualization:** Zigui Chen.

**Writing – original draft:** Zigui Chen, Rob DeSalle, Mark Schiffman, Rolando Herrero, Charles E. Wood, Julio C. Ruiz, Gary M. Clifford, Paul K. S. Chan, Robert D. Burk.

**Writing – review & editing:** Zigui Chen, Robert D. Burk.

## References

1. Bernard HU, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology*. 2010; 401(1):70–9. Epub 2010/03/09. <https://doi.org/10.1016/j.virol.2010.02.002> PMID: 20206957
2. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. *Virology*. 2004; 324(1):17–27. <https://doi.org/10.1016/j.virol.2004.03.033> PMID: 15183049
3. Van Doorslaer K, Chen Z, Bernard HU, Chan PKS, DeSalle R, Dillner J, et al. ICTV Virus Taxonomy Profile: Papillomaviridae. *J Gen Virol*. 2018. Epub 2018/06/22.
4. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Biological agents. Volume 100 B. A review of human carcinogens. IARC monographs on the evaluation of carcinogenic risks to humans / World Health Organization, International Agency for Research on Cancer. 2012; 100(Pt B):1–441.
5. Schiffman M, Doorbar J, Wentzensen N, de Sanjose S, Fakhry C, Monk BJ, et al. Carcinogenic human papillomavirus infection. *Nat Rev Dis Primers*. 2016; 2:16086. Epub 2016/12/03. <https://doi.org/10.1038/nrdp.2016.86> PMID: 27905473
6. Wood CE, Chen Z, Cline JM, Miller BE, Burk RD. Characterization and experimental transmission of an oncogenic papillomavirus in female macaques. *J Virol*. 2007; 81(12):6339–45. Epub 2007/04/13. <https://doi.org/10.1128/JVI.00233-07> PMID: 17428865
7. Van Doorslaer K, Burk RD. Evolution of human papillomavirus carcinogenicity. *Adv Virus Res*. 2010; 77:41–62. Epub 2010/10/19. <https://doi.org/10.1016/B978-0-12-385034-8.00002-8> PMID: 20951869
8. Burk RD, Ho GY, Beardsley L, Lempa M, Peters M, Bierman R. Sexual behavior and partner characteristics are the predominant risk factors for genital human papillomavirus infection in young women. *J Infect Dis*. 1996; 174(4):679–89. PMID: 8843203
9. Herrero R, Castle PE, Schiffman M, Bratti MC, Hildesheim A, Morales J, et al. Epidemiologic profile of type-specific human papillomavirus infection and cervical neoplasia in Guanacaste, Costa Rica. *J Infect Dis*. 2005; 191(11):1796–807. <https://doi.org/10.1086/428850> PMID: 15871111
10. Li N, Franceschi S, Howell-Jones R, Snijders PJ, Clifford GM. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int J Cancer*. 2011; 128(4):927–35. Epub 2010/05/18. <https://doi.org/10.1002/ijc.25396> PMID: 20473886

11. Smith JS, Lindsay L, Hoots B, Keys J, Franceschi S, Winer R, et al. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. *Int J Cancer*. 2007; 121(3):621–32. Epub 2007/04/04. <https://doi.org/10.1002/ijc.22527> PMID: [17405118](https://pubmed.ncbi.nlm.nih.gov/17405118/)
12. Kjaer SK, Munk C, Junge J, Iftner T. Carcinogenic HPV prevalence and age-specific type distribution in 40,382 women with normal cervical cytology, ASCUS/LSIL, HSIL, or cervical cancer: what is the potential for prevention? *Cancer Causes Control*. 2014; 25(2):179–89. Epub 2013/11/19. <https://doi.org/10.1007/s10552-013-0320-z> PMID: [24242002](https://pubmed.ncbi.nlm.nih.gov/24242002/)
13. Small W Jr., Bacon MA, Bajaj A, Chuang LT, Fisher BJ, Harkenrider MM, et al. Cervical cancer: A global health crisis. *Cancer*. 2017; 123(13):2404–12. Epub 2017/05/04. <https://doi.org/10.1002/ncr.30667> PMID: [28464289](https://pubmed.ncbi.nlm.nih.gov/28464289/)
14. Antonsson A, Forslund O, Ekberg H, Sterner G, Hansson BG. The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses. *J Virol*. 2000; 74(24):11636–41. PMID: [11090162](https://pubmed.ncbi.nlm.nih.gov/11090162/)
15. Woolhouse ME, Webster JP, Domingo E, Charlesworth B, Levin BR. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat Genet*. 2002; 32(4):569–77. Epub 2002/11/29. <https://doi.org/10.1038/ng1202-569> PMID: [12457190](https://pubmed.ncbi.nlm.nih.gov/12457190/)
16. Hafner MS, Nadler SA. Phylogenetic trees support the coevolution of parasites and their hosts. *Nature*. 1988; 332(6161):258–9. <https://doi.org/10.1038/332258a0> PMID: [3347269](https://pubmed.ncbi.nlm.nih.gov/3347269/)
17. Rector A, Lemey P, Tachezy R, Mostmans S, Ghim SJ, Van Doorslaer K, et al. Ancient papillomavirus-host co-speciation in Felidae. *Genome Biol*. 2007; 8(4):R57. <https://doi.org/10.1186/gb-2007-8-4-r57> PMID: [17430578](https://pubmed.ncbi.nlm.nih.gov/17430578/)
18. Bernard HU, Calleja-Macias IE, Dunn ST. Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer*. 2006; 118(5):1071–6. <https://doi.org/10.1002/ijc.21655> PMID: [16331617](https://pubmed.ncbi.nlm.nih.gov/16331617/)
19. Chen Z, Ho WCS, Boon SS, Law PTY, Chan MCW, DeSalle R, et al. Ancient Evolution and Dispersion of Human Papillomavirus 58 Variants. *J Virol*. 2017; 91(21). Epub 2017/08/11.
20. Pimenoff VN, de Oliveira CM, Bravo IG. Transmission between Archaic and Modern Human Ancestors during the Evolution of the Oncogenic Human Papillomavirus 16. *Mol Biol Evol*. 2017; 34(1):4–19. <https://doi.org/10.1093/molbev/msw214> PMID: [28025273](https://pubmed.ncbi.nlm.nih.gov/28025273/)
21. Gottschling M, Stamatakis A, Nindl I, Stockfleth E, Alonso A, Bravo IG. Multiple evolutionary mechanisms drive papillomavirus diversification. *Mol Biol Evol*. 2007; 24(5):1242–58. <https://doi.org/10.1093/molbev/msm039> PMID: [17344207](https://pubmed.ncbi.nlm.nih.gov/17344207/)
22. Shah SD, Doorbar J, Goldstein RA. Analysis of host-parasite incongruence in papillomavirus evolution using importance sampling. *Mol Biol Evol*. 2010; 27(6):1301–14. Epub 2010/01/23. <https://doi.org/10.1093/molbev/msq015> PMID: [20093429](https://pubmed.ncbi.nlm.nih.gov/20093429/)
23. Burk RD, Chen Z, Van Doorslaer K. Human papillomaviruses: genetic basis of carcinogenicity. *Public Health Genomics*. 2009; 12(5–6):281–90. Epub 2009/08/18. <https://doi.org/10.1159/000214919> PMID: [19684441](https://pubmed.ncbi.nlm.nih.gov/19684441/)
24. Silvestre RV, de Souza AJ, Junior EC, Silva AK, de Mello WA, Nunes MR, et al. First New World Primate Papillomavirus Identification in the Atlantic Forest, Brazil: *Alouatta guariba papillomavirus 1*. *Genome announcements*. 2016; 4(4). Epub 2016/08/20.
25. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*. 2007; 316(5822):222–34. <https://doi.org/10.1126/science.1139247> PMID: [17431167](https://pubmed.ncbi.nlm.nih.gov/17431167/)
26. Perez SI, Tejedor MF, Novo NM, Aristide L. Divergence Times and the Evolutionary Radiation of New World Monkeys (Platyrrhini, Primates): An Analysis of Fossil and Molecular Data. *PLoS ONE*. 2013; 8(6):e68029. <https://doi.org/10.1371/journal.pone.0068029> PMID: [23826358](https://pubmed.ncbi.nlm.nih.gov/23826358/)
27. Chen Z, van Doorslaer K, DeSalle R, Wood CE, Kaplan JR, Wagner JD, et al. Genomic diversity and interspecies host infection of alpha12 *Macaca fascicularis* papillomaviruses (MfPVs). *Virology*. 2009; 393(2):304–10. Epub 2009/09/01. <https://doi.org/10.1016/j.virol.2009.07.012> PMID: [19716580](https://pubmed.ncbi.nlm.nih.gov/19716580/)
28. Joh J, Hopper K, Van Doorslaer K, Sundberg JP, Jenson AB, Ghim SJ. *Macaca fascicularis* papillomavirus type 1: a non-human primate betapapillomavirus causing rapidly progressive hand and foot papillomatosis. *J Gen Virol*. 2009; 90(Pt 4):987–94. Epub 2009/03/07. <https://doi.org/10.1099/vir.0.006544-0> PMID: [19264664](https://pubmed.ncbi.nlm.nih.gov/19264664/)
29. Wood CE, Tannehill-Gregg SH, Chen Z, Doorslaer K, Nelson DR, Cline JM, et al. Novel betapapillomavirus associated with hand and foot papillomas in a cynomolgus macaque. *Vet Pathol*. 2011; 48(3):731–6. Epub 2010/10/06. <https://doi.org/10.1177/0300985810383875> PMID: [20921322](https://pubmed.ncbi.nlm.nih.gov/20921322/)
30. Antonsson A, Hansson BG. Healthy skin of many animal species harbors papillomaviruses which are closely related to their human counterparts. *J Virol*. 2002; 76(24):12537–42. Epub 2002/11/20. <https://doi.org/10.1128/JVI.76.24.12537-12542.2002> PMID: [12438579](https://pubmed.ncbi.nlm.nih.gov/12438579/)

31. Chan SY, Bernard HU, Ratterree M, Birkebak TA, Faras AJ, Ostrow RS. Genomic diversity and evolution of papillomaviruses in rhesus monkeys. *J Virol.* 1997; 71(7):4938–43. PMID: [9188556](#)
32. Van Ranst M, Fuse A, Sobis H, De Meurichy W, Syrjanen SM, Billiau A, et al. A papillomavirus related to HPV type 13 in oral focal epithelial hyperplasia in the pygmy chimpanzee. *J Oral Pathol Med.* 1991; 20(7):325–31. Epub 1991/08/01. PMID: [1654423](#)
33. Chen Z, Wood CE, Abee CR, Burk RD. Complete Genome Sequences of Three Novel Saimiri sciureus Papillomavirus Types Isolated from the Cervicovaginal Region of Squirrel Monkeys. *Genome announcements.* 2018; 6(1). Epub 2018/01/06.
34. Narechania A, Chen Z, DeSalle R, Burk RD. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J Virol.* 2005; 79(24):15503–10. Epub 2005/11/25. <https://doi.org/10.1128/JVI.79.24.15503-15510.2005> PMID: [16306621](#)
35. de Vienne DM, Refregier G, Lopez-Villavicencio M, Tellier A, Hood ME, Giraud T. Cospeciation vs host-shift speciation: methods for testing, evidence from natural associations and relation to coevolution. *The New phytologist.* 2013; 198(2):347–85. <https://doi.org/10.1111/nph.12150> PMID: [23437795](#)
36. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral ecology.* 2001; 26(1):32–46.
37. Bravo IG, Felez-Sanchez M. Papillomaviruses: Viral evolution, cancer and evolutionary medicine. *Evol Med Public Health.* 2015; 2015(1):32–51. Epub 2015/01/31. <https://doi.org/10.1093/emph/eov003> PMID: [25634317](#)
38. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science.* 2007; 316(5822):222–34. <https://doi.org/10.1126/science.1139247> PMID: [17431167](#)
39. Stevens NJ, Seiffert ER, O'Connor PM, Roberts EM, Schmitz MD, Krause C, et al. Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature.* 2013; 497(7451):611–4. Epub 2013/05/17. <https://doi.org/10.1038/nature12161> PMID: [23676680](#)
40. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature.* 2006; 441(7097):1103–8. <https://doi.org/10.1038/nature04789> PMID: [16710306](#)
41. Li N, Franceschi S, Howell-Jones R, Snijders PJ, Clifford GM. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. *Int J Cancer.* 2011; 128(4):927–35. Epub 2010/05/18. <https://doi.org/10.1002/ijc.25396> PMID: [20473886](#)
42. Herfs M, Soong TR, Delvenne P, Crum CP. Deciphering the Multifactorial Susceptibility of Mucosal Junction Cells to HPV Infection and Related Carcinogenesis. *Viruses.* 2017; 9(4). Epub 2017/04/21.
43. Burk RD, Harari A, Chen Z. Human papillomavirus genome variants. *Virology.* 2013; 445(1–2):232–43. Epub 2013/09/04. <https://doi.org/10.1016/j.virol.2013.07.018> PMID: [23998342](#)
44. Yamada T, Wheeler CM, Halpern AL, Stewart AC, Hildesheim A, Jenison SA. Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. *J Virol.* 1995; 69(12):7743–53. PMID: [7494284](#)
45. Stringer C. The origin and evolution of Homo sapiens. *Philos Trans R Soc Lond B Biol Sci.* 2016; 371(1698).
46. Langgut D, Almogi-Labin A, Bar-Matthews M, Pickarski N, Weinstein-Evron M. Evidence for a humid interval at approximately 56–44 ka in the Levant and its potential link to modern humans dispersal out of Africa. *J Hum Evol.* 2018. Epub 2018/09/05.
47. Lamb HF, Bates CR, Bryant CL, Davies SJ, Huws DG, Marshall MH, et al. 150,000-year palaeoclimate record from northern Ethiopia supports early, multiple dispersals of modern humans from Africa. *Scientific reports.* 2018; 8(1):1077. Epub 2018/01/20. <https://doi.org/10.1038/s41598-018-19601-w> PMID: [29348464](#)
48. Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature.* 2014; 505(7481):43–9. <https://doi.org/10.1038/nature12886> PMID: [24352235](#)
49. Buck CB, Van Doorslaer K, Peretti A, Geoghegan EM, Tisza MJ, An P, et al. The Ancient Evolutionary History of Polyomaviruses. *PLoS Pathog.* 2016; 12(4):e1005574. <https://doi.org/10.1371/journal.ppat.1005574> PMID: [27093155](#)
50. McGeoch DJ, Rixon FJ, Davison AJ. Topics in herpesvirus genomics and evolution. *Virus Res.* 2006; 117(1):90–104. Epub 2006/02/24. <https://doi.org/10.1016/j.virusres.2006.01.002> PMID: [16490275](#)
51. Niewiadomska AM, Gifford RJ. The extraordinary evolutionary history of the reticuloendotheliosis viruses. *PLoS Biol.* 2013; 11(8):e1001642. Epub 2013/09/10. <https://doi.org/10.1371/journal.pbio.1001642> PMID: [24013706](#)

52. Gottschling M, Goker M, Stamatakis A, Bininda-Emonds OR, Nindl I, Bravo IG. Quantifying the phylogenetic forces driving papillomavirus evolution. *Mol Biol Evol*. 2011; 28(7):2101–13. Epub 2011/02/03. <https://doi.org/10.1093/molbev/msr030> PMID: 21285031
53. Mirabello L, Yeager M, Cullen M, Boland JF, Chen Z, Wentzensen N, et al. HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women. *J Natl Cancer Inst*. 2016; 108(9). Epub 2016/05/01.
54. Schiffman M, Rodriguez AC, Chen Z, Wacholder S, Herrero R, Hildesheim A, et al. A population-based prospective study of carcinogenic human papillomavirus variant lineages, viral persistence, and cervical neoplasia. *Cancer Res*. 2010; 70(8):3159–69. Epub 2010/04/01. <https://doi.org/10.1158/0008-5472.CAN-09-4179> PMID: 20354192
55. Xi LF, Koutsky LA, Hildesheim A, Galloway DA, Wheeler CM, Winer RL, et al. Risk for high-grade cervical intraepithelial neoplasia associated with variants of human papillomavirus types 16 and 18. *Cancer Epidemiol Biomarkers Prev*. 2007; 16(1):4–10. <https://doi.org/10.1158/1055-9965.EPI-06-0670> PMID: 17220325
56. Arsuaga JL, Martinez I, Arnold LJ, Aranburu A, Gracia-Tellez A, Sharp WD, et al. Neandertal roots: Cranial and chronological evidence from Sima de los Huesos. *Science*. 2014; 344(6190):1358–63. <https://doi.org/10.1126/science.1253958> PMID: 24948730
57. Kuhlwilm M, Gronau I, Hubisz MJ, de Filippo C, Prado-Martinez J, Kircher M, et al. Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature*. 2016; 530(7591):429–33. <https://doi.org/10.1038/nature16544> PMID: 26886800
58. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. *Science*. 2010; 328(5979):710–22. Epub 2010/05/08. <https://doi.org/10.1126/science.1188021> PMID: 20448178
59. Meyer M, Arsuaga JL, de Filippo C, Nagel S, Aximu-Petri A, Nickel B, et al. Nuclear DNA sequences from the Middle Pleistocene Sima de los Huesos hominins. *Nature*. 2016; 531(7595):504–7. <https://doi.org/10.1038/nature17405> PMID: 26976447
60. Vernot B, Tucci S, Kelso J, Schraiber JG, Wolf AB, Gittelman RM, et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science*. 2016; 352(6282):235–9. Epub 2016/03/19. <https://doi.org/10.1126/science.aad9416> PMID: 26989198
61. Skoglund P, Jakobsson M. Archaic human ancestry in East Asia. *Proc Natl Acad Sci U S A*. 2011; 108(45):18301–6. Epub 2011/11/02. <https://doi.org/10.1073/pnas.1108181108> PMID: 22042846
62. Hang D, Yin Y, Han J, Jiang J, Ma H, Xie S, et al. Analysis of human papillomavirus 16 variants and risk for cervical cancer in Chinese population. *Virology*. 2016; 488:156–61. Epub 2015/12/10. <https://doi.org/10.1016/j.virol.2015.11.016> PMID: 26650690
63. Cornet I, Gheit T, Franceschi S, Vignat J, Burk RD, Sylla BS, et al. Human papillomavirus type 16 genetic variants: phylogeny and classification based on E6 and LCR. *J Virol*. 2012; 86(12):6855–61. Epub 2012/04/12. <https://doi.org/10.1128/JVI.00483-12> PMID: 22491459
64. Reich D, Green RE, Kircher M, Krause J, Patterson N, Durand EY, et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 2010; 468(7327):1053–60. Epub 2010/12/24. <https://doi.org/10.1038/nature09710> PMID: 21179161
65. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015; 526(7571):68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245
66. Raghavan M, Steinrucken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, et al. POPULATION GENETICS. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science*. 2015; 349(6250):aab3884. Epub 2015/07/23. <https://doi.org/10.1126/science.aab3884> PMID: 26198033
67. Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo JM, et al. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell*. 2012; 150(3):457–69. Epub 2012/07/31. <https://doi.org/10.1016/j.cell.2012.07.009> PMID: 22840920
68. Veeramah KR, Hammer MF. The impact of whole-genome sequencing on the reconstruction of human population history. *Nat Rev Genet*. 2014; 15(3):149–62. Epub 2014/02/05. <https://doi.org/10.1038/nrg3625> PMID: 24492235
69. Vattathil S, Akey JM. Small Amounts of Archaic Admixture Provide Big Insights into Human History. *Cell*. 2015; 163(2):281–4. Epub 2015/10/10. <https://doi.org/10.1016/j.cell.2015.09.042> PMID: 26451479
70. Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science*. 2011; 334(6052):89–94. Epub 2011/08/27. <https://doi.org/10.1126/science.1209202> PMID: 21868630

71. Chen D, Enroth S, Liu H, Sun Y, Wang H, Yu M, et al. Pooled analysis of genome-wide association studies of cervical intraepithelial neoplasia 3 (CIN3) identifies a new susceptibility locus. *Oncotarget*. 2016; 7(27):42216–24. Epub 2016/06/11. <https://doi.org/10.18632/oncotarget.9916> PMID: 27285765
72. Safaeian M, Johnson LG, Yu K, Wang SS, Gravitt PE, Hansen JA, et al. Human Leukocyte Antigen Class I and II Alleles and Cervical Adenocarcinoma. *Front Oncol*. 2014; 4:119. Epub 2014/07/06. <https://doi.org/10.3389/fonc.2014.00119> PMID: 24995157
73. Madeleine MM, Johnson LG, Smith AG, Hansen JA, Nisperos BB, Li S, et al. Comprehensive analysis of HLA-A, HLA-B, HLA-C, HLA-DRB1, and HLA-DQB1 loci and squamous cell cervical cancer risk. *Cancer Res*. 2008; 68(9):3532–9. Epub 2008/05/03. <https://doi.org/10.1158/0008-5472.CAN-07-6471> PMID: 18451182
74. Racimo F, Sankararaman S, Nielsen R, Huerta-Sanchez E. Evidence for archaic adaptive introgression in humans. *Nat Rev Genet*. 2015; 16(6):359–71. Epub 2015/05/13. <https://doi.org/10.1038/nrg3936> PMID: 25963373
75. Galway-Witham J, Stringer C. How did Homo sapiens evolve? *Science*. 2018; 360(6395):1296–8. Epub 2018/06/23. <https://doi.org/10.1126/science.aat6659> PMID: 29930123
76. Egawa N, Egawa K, Griffin H, Doorbar J. Human Papillomaviruses; Epithelial Tropisms, and the Development of Neoplasia. *Viruses*. 2015; 7(7):3863–90. Epub 2015/07/21. <https://doi.org/10.3390/v7072802> PMID: 26193301
77. Castle PE, Schiffman M, Gravitt PE, Kendall H, Fishman S, Dong H, et al. Comparisons of HPV DNA detection by MY09/11 PCR methods. *J Med Virol*. 2002; 68(3):417–23. <https://doi.org/10.1002/jmv.10220> PMID: 12226831
78. Forslund O, Antonsson A, Nordin P, Stenquist B, Hansson BG. A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin. *J Gen Virol*. 1999; 80 (Pt 9):2437–43.
79. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012; 28(12):1647–9. <https://doi.org/10.1093/bioinformatics/bts199> PMID: 22543367
80. Fonseca AJ, Taeko D, Chaves TA, Amorim LD, Murari RS, Miranda AE, et al. HPV Infection and Cervical Screening in Socially Isolated Indigenous Women Inhabitants of the Amazonian Rainforest. *PLoS ONE*. 2015; 10(7):e0133635. Epub 2015/07/25. <https://doi.org/10.1371/journal.pone.0133635> PMID: 26207895
81. Agalliu I, Gapstur S, Chen Z, Wang T, Anderson RL, Teras L, et al. Associations of Oral alpha-, beta-, and gamma-Human Papillomavirus Types With Risk of Incident Head and Neck Cancer. *JAMA Oncol*. 2016. Epub 2016/01/23.
82. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. 2016; 102:3–11. Epub 2016/03/26. <https://doi.org/10.1016/j.ymeth.2016.02.020> PMID: 27012178
83. Smith B, Chen Z, Reimers L, van Doorslaer K, Schiffman M, Desalle R, et al. Sequence imputation of HPV16 genomes for genetic association studies. *PLoS ONE*. 2011; 6(6):e21375. Epub 2011/07/07. <https://doi.org/10.1371/journal.pone.0021375> PMID: 21731721
84. Chen Z, Terai M, Fu L, Herrero R, DeSalle R, Burk RD. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. *J Virol*. 2005; 79(11):7014–23. Epub 2005/05/14. <https://doi.org/10.1128/JVI.79.11.7014-7023.2005> PMID: 15890941
85. Castle PE, Shaber R, LaMere BJ, Kinney W, Fetterma B, Poitras N, et al. Human papillomavirus (HPV) genotypes in women with cervical precancer and cancer at Kaiser Permanente Northern California. *Cancer Epidemiol Biomarkers Prev*. 2011; 20(5):946–53. <https://doi.org/10.1158/1055-9965.EPI-10-1267> PMID: 21415357
86. Chen Z, Schiffman M, Herrero R, Desalle R, Anastos K, Segondy M, et al. Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. *PLoS ONE*. 2011; 6(5):e20183. Epub 2011/06/16. <https://doi.org/10.1371/journal.pone.0020183> PMID: 21673791
87. Cullen M, Boland JF, Schiffman M, Zhang X, Wentzensen N, Yang Q, et al. Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection. *Papillomavirus Res*. 2015; 1:3–11. Epub 2015/12/09. <https://doi.org/10.1016/j.pvr.2015.05.004> PMID: 26645052
88. Kennedy IM, Haddow JK, Clements JB. A negative regulatory element in the human papillomavirus type 16 genome acts at the level of late mRNA stability. *J Virol*. 1991; 65(4):2093–7. Epub 1991/04/01. PMID: 1848319

89. Flores ER, Allen-Hoffmann BL, Lee D, Sattler CA, Lambert PF. Establishment of the human papillomavirus type 16 (HPV-16) life cycle in an immortalized human foreskin keratinocyte cell line. *Virology*. 1999; 262(2):344–54. <https://doi.org/10.1006/viro.1999.9868> PMID: 10502513
90. Kirnbauer R, Taub J, Greenstone H, Roden R, Durst M, Gissmann L, et al. Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles. *J Virol*. 1993; 67(12):6929–36. PMID: 8230414
91. Lurchachaiwong W, Junyangdikul P, Payungporn S, Chansaenroj J, Sampathanukul P, Tresukosol D, et al. Entire genome characterization of human papillomavirus type 16 from infected Thai women with different cytological findings. *Virus Genes*. 2009; 39(1):30–8. Epub 2009/05/05. <https://doi.org/10.1007/s11262-009-0363-0> PMID: 19412733
92. Seedorf K, Krammer G, Durst M, Suhai S, Rowekamp WG. Human papillomavirus type 16 DNA sequence. *Virology*. 1985; 145(1):181–5. Epub 1985/08/01. PMID: 2990099
93. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004; 32(5):1792–7. Epub 2004/03/23. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
94. Katoh K, Toh H. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*. 2010; 26(15):1899–900. Epub 2010/04/30. <https://doi.org/10.1093/bioinformatics/btq224> PMID: 20427515
95. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22(21):2688–90. Epub 2006/08/25. <https://doi.org/10.1093/bioinformatics/btl446> PMID: 16928733
96. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 2003; 52(5):696–704. Epub 2003/10/08. PMID: 14530136
97. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003; 19(12):1572–4. PMID: 12912839
98. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998; 14(9):817–8. PMID: 9918953
99. Miller MA, Pfeiffer W, Schwartz T, editors. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE), 2010; 2010: IEEE.
100. Didelot-Rousseau MN, Nagot N, Costes-Martineau V, Valles X, Ouedraogo A, Konate I, et al. Human papillomavirus genotype distribution and cervical squamous intraepithelial lesions among high-risk women with and without HIV-1 infection in Burkina Faso. *Br J Cancer*. 2006; 95(3):355–62. <https://doi.org/10.1038/sj.bjc.6603252> PMID: 16832413
101. Okolo C, Franceschi S, Adewole I, Thomas JO, Follen M, Snijders PJ, et al. Human papillomavirus infection in women with and without cervical cancer in Ibadan, Nigeria. *Infectious agents and cancer*. 2010; 5(1):24. Epub 2010/12/07. <https://doi.org/10.1186/1750-9378-5-24> PMID: 21129194
102. Singh DK, Anastos K, Hoover DR, Burk RD, Shi Q, Ngendahayo L, et al. Human Papillomavirus Infection and Cervical Cytology in HIV-Infected and HIV-Uninfected Rwandan Women. *J Infect Dis*. 2009; 199(12):1851–61. Epub 2009/05/14. <https://doi.org/10.1086/599123> PMID: 19435429
103. Tornesello ML, Buonaguro FM, Meglio A, Buonaguro L, Beth-Giraldo E, Giraldo G. Sequence variations and viral genomic state of human papillomavirus type 16 in penile carcinomas from Ugandan patients. *J Gen Virol*. 1997; 78 (Pt 9):2199–208.
104. Sahasrabudde VV, Mwanahamuntu MH, Vermund SH, Huh WK, Lyon MD, Stringer JS, et al. Prevalence and distribution of HPV genotypes among HIV-infected women in Zambia. *Br J Cancer*. 2007; 96(9):1480–3. <https://doi.org/10.1038/sj.bjc.6603737> PMID: 17437020
105. Chan PK, Lam CW, Cheung TH, Li WW, Lo KW, Chan MY, et al. Human papillomavirus type 16 intra-typic variant infection and risk for cervical neoplasia in southern China. *J Infect Dis*. 2002; 186(5):696–700. Epub 2002/08/27. <https://doi.org/10.1086/342048> PMID: 12195358
106. Wu Y, Liu B, Lin W, Xu Y, Li L, Zhang Y, et al. HPV16 E6 variants and HLA class II polymorphism among Chinese women with cervical cancer. *J Med Virol*. 2007; 79(4):439–46. <https://doi.org/10.1002/jmv.20785> PMID: 17311339
107. Liaw K-L, Hsing A, Chen C-J, Schiffman M, Zhang T, Hsieh C-Y, et al. Human papillomavirus and cervical neoplasia: a case-control study in Taiwan. *Int J Cancer*. 1995; 62:565–71. PMID: 7665227
108. Pillai MR, Hariharan R, Babu JM, Lakshmi S, Chiplunkar SV, Patkar M, et al. Molecular variants of HPV-16 associated with cervical cancer in Indian population. *Int J Cancer*. 2009. Epub 2009/04/10.
109. Bhattacharjee B, Sengupta S. HPV16 E2 gene disruption and polymorphisms of E2 and LCR: some significant associations with cervical cancer in Indian women. *Gynecol Oncol*. 2006; 100(2):372–8. <https://doi.org/10.1016/j.ygyno.2005.09.016> PMID: 16246404

110. Matsumoto K, Yasugi T, Nakagawa S, Okubo M, Hirata R, Maeda H, et al. Human papillomavirus type 16 E6 variants and HLA class II alleles among Japanese women with cervical cancer. *Int J Cancer*. 2003; 106(6):919–22. Epub 2003/08/15. <https://doi.org/10.1002/ijc.11332> PMID: 12918070
111. Choi BS, Kim SS, Yun H, Jang DH, Lee JS. Distinctive distribution of HPV16 E6 D25E and E7 N29S intratypic Asian variants in Korean commercial sex workers. *J Med Virol*. 2007; 79(4):426–30. <https://doi.org/10.1002/jmv.20826> PMID: 17311337
112. Wongworapat K, Keawwichit R, Sirirojn B, Dokuta S, Ruangyuttikarn C, Sriplienchan S, et al. Detection of human papillomavirus from self-collected vaginal samples of women in Chiang Mai, Thailand. *Sex Transm Dis*. 2008; 35(2):172–3. Epub 2008/01/25. <https://doi.org/10.1097/OLQ.0b013e318158af65> PMID: 18216725
113. Marks M, Gravitt PE, Gupta SB, Liaw KL, Kim E, Tadesse A, et al. The association of hormonal contraceptive use and HPV prevalence. *Int J Cancer*. 2011. Epub 2010/08/25.
114. Hoffmann M, Lohrey C, Hunziker A, Kahn T, Schwarz E. Human papillomavirus type 16 E6 and E7 genotypes in head-and-neck carcinomas. *Oral oncology*. 2004; 40(5):520–4. Epub 2004/03/10. <https://doi.org/10.1016/j.oraloncology.2003.10.011> PMID: 15006625
115. Tanzi E, Amendola A, Bianchi S, Fasolo MM, Beretta R, Pariani E, et al. Human papillomavirus genotypes and phylogenetic analysis of HPV-16 variants in HIV-1 infected subjects in Italy. *Vaccine*. 2009; 27 Suppl 1:A17–23. Epub 2009/06/02.
116. Cento V, Ciccozzi M, Ronga L, Perno CF, Ciotti M. Genetic diversity of human papillomavirus type 16 E6, E7, and L1 genes in Italian women with different grades of cervical lesions. *J Med Virol*. 2009; 81(9):1627–34. Epub 2009/07/25. <https://doi.org/10.1002/jmv.21552> PMID: 19626616
117. Tornesello ML, Duraturo ML, Giorgi-Rossi P, Sansone M, Piccoli R, Buonaguro L, et al. Human papillomavirus (HPV) genotypes and HPV16 variants in human immunodeficiency virus-positive Italian women. *J Gen Virol*. 2008; 89(Pt 6):1380–9. Epub 2008/05/14. <https://doi.org/10.1099/vir.0.83553-0> PMID: 18474553
118. Tornesello ML, Duraturo ML, Salatiello I, Buonaguro L, Losito S, Botti G, et al. Analysis of human papillomavirus type-16 variants in Italian women with cervical intraepithelial neoplasia and cervical cancer. *Journal of medical virology*. 2004; 74(1):117–26. Epub 2004/07/20. <https://doi.org/10.1002/jmv.20154> PMID: 15258977
119. de Boer MA, Peters LA, Aziz MF, Siregar B, Cornain S, Vrede MA, et al. Human papillomavirus type 16 E6, E7, and L1 variants in cervical cancer in Indonesia, Suriname, and The Netherlands. *Gynecol Oncol*. 2004; 94(2):488–94. Epub 2004/08/07. <https://doi.org/10.1016/j.ygyno.2004.05.037> PMID: 15297193
120. Bontkes HJ, van Duin M, de Grujil TD, Duggan-Keen MF, Walboomers JM, Stukart MJ, et al. HPV 16 infection and progression of cervical intra-epithelial neoplasia: analysis of HLA polymorphism and HPV 16 E6 sequence variants. *International journal of cancer Journal international du cancer*. 1998; 78(2):166–71. Epub 1998/10/01. PMID: 9754647
121. Pista A, Oliveira A, Barateiro A, Costa H, Verdasca N, Paixao MT. Molecular variants of human papillomavirus type 16 and 18 and risk for cervical neoplasia in Portugal. *J Med Virol*. 2007; 79(12):1889–97. <https://doi.org/10.1002/jmv.21002> PMID: 17935194
122. Hu X, Pang T, Guo Z, Mazurenko N, Kissel'jov F, Ponten J, et al. HPV16 E6 gene variations in invasive cervical squamous cell carcinoma and cancer in situ from Russian patients. *Br J Cancer*. 2001; 84(6):791–5. <https://doi.org/10.1054/bjoc.2000.1619> PMID: 11259093
123. Ortiz M, Torres M, Munoz L, Fernandez-Garcia E, Canals J, Cabornero AI, et al. Oncogenic human papillomavirus (HPV) type distribution and HPV type 16 E6 variants in two Spanish population groups with different levels of HPV infection risk. *J Clin Microbiol*. 2006; 44(4):1428–34. <https://doi.org/10.1128/JCM.44.4.1428-1434.2006> PMID: 16597872
124. Bible JM, Mant C, Best JM, Kell B, Starkey WG, Shanti Raju K, et al. Cervical lesions are associated with human papillomavirus type 16 intratypic variants that have high transcriptional activity and increased usage of common mammalian codons. *J Gen Virol*. 2000; 81(Pt 6):1517–27. <https://doi.org/10.1099/0022-1317-81-6-1517> PMID: 10811935
125. Alencar TR, Cerqueira DM, da Cruz MR, Wyant PS, Ramalho ED, Martins CR. New HPV-16 European and non-European variants in Central Brazil. *Virus Genes*. 2007; 35(1):1–4. <https://doi.org/10.1007/s11262-006-0040-5> PMID: 17048111
126. de Araujo Souza PS, Maciag PC, Ribeiro KB, Petzl-Erler ML, Franco EL, Villa LL. Interaction between polymorphisms of the human leukocyte antigen and HPV-16 variants on the risk of invasive cervical cancer. *BMC Cancer*. 2008; 8:246. Epub 2008/08/30. <https://doi.org/10.1186/1471-2407-8-246> PMID: 18721466
127. Junes-Gill K, Sichero L, Maciag PC, Mello W, Noronha V, Villa LL. Human papillomavirus type 16 variants in cervical cancer from an admixed population in Brazil. *Journal of medical virology*. 2008; 80(9):1639–45. Epub 2008/07/24. <https://doi.org/10.1002/jmv.21238> PMID: 18649325

128. Berumen J, Ordonez RM, Lazcano E, Salmeron J, Galvan SC, Estrada RA, et al. Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study. *J Natl Cancer Inst.* 2001; 93(17):1325–30. PMID: [11535707](https://pubmed.ncbi.nlm.nih.gov/11535707/)
129. Calleja-Macias IE, Kalantari M, Huh J, Ortiz-Lopez R, Rojas-Martinez A, Gonzalez-Guerrero JF, et al. Genomic diversity of human papillomavirus-16, 18, 31, and 35 isolates in a Mexican population and relationship to European, African, and Native American variants. *Virology.* 2004; 319(2):315–23. <https://doi.org/10.1016/j.virol.2003.11.009> PMID: [14980491](https://pubmed.ncbi.nlm.nih.gov/14980491/)
130. Lopez-Revilla R, Pineda MA, Ortiz-Valdez J, Sanchez-Garza M, Riego L. Human papillomavirus type 16 variants in cervical intraepithelial neoplasia and invasive carcinoma in San Luis Potosi City, Mexico. *Infectious agents and cancer.* 2009; 4(1):3. Epub 2009/02/17.
131. Lizano M, De la Cruz-Hernandez E, Carrillo-Garcia A, Garcia-Carranca A, Ponce de Leon-Rosales S, Duenas-Gonzalez A, et al. Distribution of HPV16 and 18 intratypic variants in normal cytology, intraepithelial lesions, and cervical cancer in a Mexican population. *Gynecol Oncol.* 2006; 102(2):230–5. <https://doi.org/10.1016/j.ygyno.2005.12.002> PMID: [16427686](https://pubmed.ncbi.nlm.nih.gov/16427686/)
132. Watts KJ, Thompson CH, Cossart YE, Rose BR. Sequence variation and physical state of human papillomavirus type 16 cervical cancer isolates from Australia and New Caledonia. *Int J Cancer.* 2002; 97(6):868–74. Epub 2002/02/22. PMID: [11857370](https://pubmed.ncbi.nlm.nih.gov/11857370/)
133. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics.* 2010; 11:538. <https://doi.org/10.1186/1471-2105-11-538> PMID: [21034504](https://pubmed.ncbi.nlm.nih.gov/21034504/)
134. Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *Isme J.* 2010; 4(1):17–27. Epub 2009/08/28. <https://doi.org/10.1038/ismej.2009.97> PMID: [19710709](https://pubmed.ncbi.nlm.nih.gov/19710709/)
135. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007; 7(1):214.
136. Stringer CB, Barnes I. Deciphering the Denisovans. *Proc Natl Acad Sci U S A.* 2015; 112(51):15542–3. <https://doi.org/10.1073/pnas.1522477112> PMID: [26668361](https://pubmed.ncbi.nlm.nih.gov/26668361/)
137. Liu W, Martinon-Torres M, Cai YJ, Xing S, Tong HW, Pei SW, et al. The earliest unequivocally modern humans in southern China. *Nature.* 2015; 526(7575):696–9. <https://doi.org/10.1038/nature15696> PMID: [26466566](https://pubmed.ncbi.nlm.nih.gov/26466566/)
138. Green RE, Krause J, Ptak SE, Briggs AW, Ronan MT, Simons JF, et al. Analysis of one million base pairs of Neanderthal DNA. *Nature.* 2006; 444(7117):330–6. Epub 2006/11/17. <https://doi.org/10.1038/nature05336> PMID: [17108958](https://pubmed.ncbi.nlm.nih.gov/17108958/)
139. Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, et al. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature.* 2014; 514(7523):445–9. Epub 2014/10/25. <https://doi.org/10.1038/nature13810> PMID: [25341783](https://pubmed.ncbi.nlm.nih.gov/25341783/)
140. Ingman M, Kaessmann H, Paabo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature.* 2000; 408(6813):708–13. Epub 2000/12/29. <https://doi.org/10.1038/35047064> PMID: [11130070](https://pubmed.ncbi.nlm.nih.gov/11130070/)
141. Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J Mol Evol.* 2006; 62(5):551–63. Epub 2006/03/25. <https://doi.org/10.1007/s00239-005-0221-1> PMID: [16557338](https://pubmed.ncbi.nlm.nih.gov/16557338/)
142. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol.* 2012; 29(9):2157–67. <https://doi.org/10.1093/molbev/mss084> PMID: [22403239](https://pubmed.ncbi.nlm.nih.gov/22403239/)