Non-human Primate Papillomaviruses Share Similar Evolutionary Histories and Niche Adaptation as the Human Counterparts

Zigui Chen1,2*, Teng Long†, Po Yee Wong1, Wendy C. S. Ho1, Robert D. Burk3 and Paul K. S. Chan1,2*

1 Department of Microbiology, The Chinese University of Hong Kong, Hong Kong, China; 2 Centre for Emerging Infectious Diseases, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, China; 3 Department of Pediatrics, Microbiology and Immunology, Epidemiology and Population Health, and Obstetrics, Gynecology and Woman’s Health, Albert Einstein College of Medicine, The Bronx, NY, United States

Given high genetic diversity of papillomaviruses (PV) and complex scenario of virus-host interaction, the genetic basis underlying the mechanisms of HPV carcinogenicity is not well understood. In an effort to evaluate the origin and evolution of PV pathogenicity, we collected paired oral, perianal, and genital swabs from a wild macaque population. Of the 117 surveyed macaques, 88 (75.2%) were positive for PV DNA in one or more sites, mostly common from genital swabs, followed by oral and perianal sites. All putative macaque PV types phylogenetically clustered into the genera Alpha-, Beta-, and Gammapapillomavirus, with a strong phylogeny-tropism association as observed in HPVs. Using a Bayesian Markov Chain Monte Carlo framework, we demonstrated ancient intra-host divergence of primate PVs in which multiple ancestors had split and adapted to specific host ecosystems at least 41 million years ago, prior to the speciation events of primate host species. Following subsequent divergence and niche adaptation, distinct but phylogenetically related PV types were transmitted to similar host ecosystems by closely related host animals when host speciation occurred, which may explain in part the origin of carcinogenicity of HPV type 16 (HPV16) and Macaca fascicularis PV type 3 (MfPV3) that evolved from a most recent common ancestor containing the determinants for cervicovaginal colonization and cervical cancer. The findings identifying evolutionary and biological relatedness between human and non-human primate PVs lay a genetic foundation for research on parasite-host interactions and carcinogenic outcomes, which will prove useful in further study of viral pathogenesis and host specificity.

STUDY IMPORTANCE

To better understand the origin and evolution of PV carcinogenicity associated with cervical cancer, we applied a combination of phylogenetic and bioinformatic analyses to investigate the genetic diversity of macaque papillomaviruses, and estimate divergence
times of human and non-human primate PVs. The majority of both human and non-human primate PVs cluster into α-, β-, and γ-PVs, sharing similar evolutionary histories and biological properties to each other. The strong phylogeny-tropism association of primate PVs indicates an important role of niche adaptation and virus-host codivergence shaping the diversity of viral genomics, host specificity, immune exposure, and pathogenic property. Understanding the evolution of the family Papilloamviridae in general and the primate papillomaviruses in specific in relevant to virus-host interactions should provide important implications for viral pathogenesis and disease prevention.

Keywords: primate papillomaviruses, niche adaptation, evolution, divergence time estimate, macaque monkey

INTRODUCTION

Papillomaviridae comprised a large family of double-stranded circular DNA viruses of about 8,000 nucleotides in size. Papillomavirus (PV) genome contains three regions, including an upstream regulatory region (URR), an early region of open reading frames (ORFs) (E1, E2, E4, E5, E6, and E7) encoding functional proteins, and a late region of ORFs encoding structural proteins (L1 and L2) (Harden and Munger, 2017). PVs infect the cutaneous and mucosal epithelia of a large spectrum of vertebrates including mammals, reptiles, and rodents even fishes (Rector and Van Ranst, 2013). To date, more than 200 types of human PVs (HPVs) and 160 types of animal PVs have been fully characterized (Van Doorslaer et al., 2018). The clinical consequence of PV infections ranges from completely asymptomatic to benign warts or even malignant neoplasia (Graham, 2017). For example, HPV type 16 (HPV16) and HPV18 contribute to over 70% of cervical cancers, the 4th most common cancer in women worldwide (Bravo et al., 2010). The majority of HPVs are classified into 3 genera: Alphapapillomavirus (α-PV), Betapapillomavirus (β-PV), and Gammapapillomavirus (γ-PV) (de Villiers et al., 2004; Bernard et al., 2010). All genital carcinogenic HPVs are members of the genus α-PV, while only a limited number of β-PV have been associated with squamous cell cancer of the skin, particularly in subjects with epidermodysplasia verruciformis (EV). The pathogenic role of γ-PV is unclear, while skin swab and oral cavity harbor abundant known and novel β- and γ-HPVs (Nindl et al., 2007; Bottalico et al., 2011; Wong et al., 2018).

Besides of carcinogenic HPVs, a large set of non-human primate PVs within the species Alphapapillomavirus 12 (α12) isolated from the cervicovaginal region of macaques have been associated with cervical cancers (Wood et al., 2007; Chen et al., 2009). The species α12 (oncogenic macaque PVs) are phylogenetic related to the species α9 (oncogenic HPV16) but distantly separated from some β- and γ-PV types that were also isolated from macaque animals (e.g., MfPV1 within the species B1), indicating that there is an evolutionary relatedness of papillomavirus carcinogenicity while they may not strictly follow the virus-host codivergence (Chen et al., 2018a). Codivergence is a process of reciprocal and cross-adaptive genetic changes, which means that the evolutionary history of pathogens should reflect that of its host counterparts in divergence times as well as phylogenetic histories (Hafner and Nadler, 1988; Woolhouse et al., 2002). For papillomaviruses, however, the prior adaptation of pathogens to various ecological niches may explain the conflicts between pathogens and hosts phylogenies (Chan et al., 1992; Bravo et al., 2010; Shah et al., 2010; Sharp and Simmonds, 2011; Chen et al., 2018a), which was also documented for other viruses, such as polyomaviruses, herpesviruses and some retrovirus genera (McGeoch et al., 2006; Niewiadomska and Gifford, 2013; Buck et al., 2016).

Papillomaviruses are ancient DNA viruses that have co-evolved with their hosts over millions of years, resulting in a high level of diversity at host specificity, tissue tropism, viral genetics, prevalence and pathogenesis (Shah et al., 2010; Van Doorslaer et al., 2017). The importance of evolutionary dynamics in driving PV divergence has been increasingly recognized recently, which largely facilitates our understanding on the genetic basis underlying the mechanisms of viral carcinogenicity (Bernard et al., 2006; Bravo et al., 2010). However, the knowledge on the origin and evolution of PV oncogenicity remain scarce. In this study, we collected oral, perianal and genital swabs from a wild macaque population (Macaca mulatta) inhabiting a protected nature reserve in Hong Kong. Genetic diversity and evolutionary relationship among macaque PVs as well as among HPVs were characterized, with divergence time estimated using Bayesian molecular clock model to trace the divergence of primate PV within niche-specific clades. The findings identifying biological and evolutionary relatedness between human and non-human primate PVs lay a genetic foundation for research on parasite-host interactions and carcinogenic outcomes, which will prove useful in further study of viral pathogenesis and host specificity.

MATERIALS AND METHODS

Macaque Sample Collection, DNA Extraction, and Papillomavirus Genotyping

Between August and December 2016, paired swab samples from oral, perianal and genital (vaginal sites from female macaques, and penial sites from male macaques) sites of 117 non-captive wild rhesus macaques (M. mulatta), including 84 females and 33 males, were collected from Kam Shan Country Park, Hong Kong, China (Chen et al., 2018b). All animals appeared healthy according to the body condition and the temperature
measured by veterinarians. Swab samples were collected in 2 ml specimen transport medium (STM) [Hank’s BSS (10×), 5% Bovine Albumin (BA), Gentamycin (4 mg/ml), Pen/Strept (50,000 µg/ml), Fungizone (1 mg/ml), and NaHCO3 7.5%] when the animals were anesthetized for the veterinary and/or contraceptive treatment and transported to the laboratory in an ice box cooler within 2 h.

The total DNA was extracted using Qiagen DNA Mini Kit (Qiagen, United States) with 200 µl STM suspension and eluted in 200 µl elution buffer (pH 8.0) for PCR-based papillomavirus genotyping. Two recently developed next-generation sequencing (NGS) amplicon assays targeting the consensus regions of PV L1 open reading frame of the full spectrum of primate papillomaviruses including α-, β-, and γ-PVs were used (Fonseca et al., 2015; Agalliu et al., 2016). Successful amplicons, with a pair of unique 12-bp barcodes introduced by forward and reverse primers, were pooled and sequenced on an Illumina Miseq (Illumina, United States) at the Weil Cornell Medicine Genomics Resources Core Facility, New York, NY, United States, using paired-end 150 bp reads.

Following the demultiplexing, the paired short reads passing the quality filter (≥Q20 and ≥50-bp) were merged into single reads using FLASH v1.2.11 (Magoc and Salzberg, 2011) and blasted against a papillomavirus reference database using UPARSE software (Edgar, 2013). An operation taxonomic unit (OTU) count table was created using a 95% identity threshold (Fonseca et al., 2015; Wong et al., 2018), with each OTU assigned for PV type based on sequence homology to the reference database: if OTUs hitting the reference database had ≥90% identities to a characterized PV type, they represented known viruses; those with 60–89% identities were regarded as “uncharacterized” types and assigned with a unique identity. A PV type was considered positive if the reads were ≥50. For each PCR amplification, multiple negative controls and random repeats were set in order to measure the false positive and minimize the potential contamination.

Characterization of Macaque Papillomavirus Complete Genomes
The complete genomes of M. mulatta papillomaviruses (MmPV2 to 7) were characterized using virus-enriched shotgun sequencing, as previously reported (Chen et al., 2018a; Long et al., 2018). Whole viral genomes were validated using type-specific PCR amplification in three or four overlapping fragments and Sanger sequencing using a primer walking strategy. ORFs were predicted using Geneious v9 and searched for homology using a BLASTX against the closely related HPV types.

Phylogenetic Analyses and Tree Construction
Phylogenetic analysis was based on the concatenated nucleotide sequence alignments of papillomavirus genes. The codon sequences from each ORF were aligned based on the corresponding amino acid sequences aligned using MUSCLE v3.8.31 (Edgar, 2004). Maximum likelihood (ML) trees were constructed using RAxML MPI v8.2.3 (Stamatakis, 2006) with optimized parameters. Data were bootstrap resampled 1,000 times. MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) with 100,000,000 cycles for the Markov chain Monte Carlo (MCMC) algorithm was used to generate Bayesian trees. A 25% 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 diagram ModelTest v3.7 (Posada and Crandall, 1998) 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 (Miller et al., 2010) was accessed to facilitate RAXML and MrBayes high-performance computation. Phylogenetic trees were visualized using FigTree v1.4.4 and Ape v5.1 (Paradis and Schliep, 2019).

Divergence Time Estimation
We used a Bayesian MCMC method implemented by BEAST v2.4.5 (Drummond and Rambaut, 2007) and the previously published PV evolutionary rates (Rector et al., 2007) to estimate the divergence time of primate PVs from their most recent common ancestors (MRCA). Separated estimations were performed for each of the genera α-, β- and γ-PV since primate PVs, taken together, do not follow strict virus-host codivergence. Three tree priors (coalescent constant population, Yule model, and coalescent Bayesian skyline) and two molecular clock models [strict mutation rate and relaxed uncorrelated lognormal distribution (UCLD)] were pre-tested, with a combination of Yule and relaxed UCLD of molecular clock model showing the lowest Akaike’s information criterion for MCMC samples (AICM) value (Baele et al., 2012). The concatenated codon sequence partitions of six ORFs (E6, E7, E1, E2, L2, and L1) with variable rates of substitution over time as previously published (Rector et al., 2007): E6 (2.39 E-08 substitutions per site per year, 95% confidence interval of 1.70 – 3.26 E-08), E7 (1.44 E-08, 0.97 – 2.00 E-08), E1 (1.76 E-08, 1.20 – 2.31 E-08), E2 (2.11 E-08, 1.52 – 2.81 E-08), L2 (2.13 E-08, 1.46 – 2.76 E-08), and L1 (1.84 E-08, 1.27 – 2.35 E-08). In order to calibrate the divergence times, we introduced three time points in the α-PV tree, with assumptions of codivergence histories between primate PVs and their hosts: (1) the split between α-PVs and Saimiri sciureus PV 1/2/3 (SscPV1/2/3, within the genus Dyooomikronpapillomavirus) at 49 (95% CI of 41 – 58) million years ago (mya) matching the divergence between Old World and New World monkey ancestors (Perez et al., 2013), (2) the separation between the species α12 and α9 at 28 (25 – 31) mya matching the speciation between hominin and macaque ancestors (Gibbs et al., 2007), and (3) the node between HPV13 and Pan paniscus PV 1 (PpPV1) at 7 (6 – 8) mya matching the split between hominin and chimpanzee ancestors (Patterson et al., 2006). For β- and γ-PV trees, the calibration time point (s) was set between macaque PVs and their closest HPV relatives. 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 were calculated using Tracer v1.6.1.
On the basis of partial L1 sequence information, a total of 72 distribution of PV genera between samples isolated from different body sites. For instance, 56 out of 65 PV-positive genital swabs (86.2%) had α-PV infection compared to 7.1% (3/42) of oral infection; whereas 90.5% (38/42) of oral infection were γ-PV compared to 32.3% (21/65) of genital infection (p < 0.001). In consensus with the tropism of body site between macaque PV genera, all 32 putative α-PV types were able to be detected in genital samples while only 2 types in oral swabs. We found 25 and 19 γ-PV types in oral and genital swabs, respectively, with 8 types in both sites (Table 1 and Supplementary Table S2). When the total number of PV infection (abundance) was considered, the tropism was also significant: genital samples had 80.3% of α-PV infections and 19.2% of γ-PV infections when compared to 4.2% of α-PV infections and 90.3% of γ-PV infections in oral cavity (p < 0.001) (Figure 1A and Table 1). The anal swabs had similar proportion of α- and γ-PV infections, inferred from the infected samples (51.4% vs. 51.4%), the unique PV types (17 vs. 18), and the PV abundance (59.7% vs. 38.9%) (Figure 1B and Table 1).

The β-PV was rare in the surveyed macaques. Only one type was detected that infected 5 animals (4 oral samples, 1 anal sample and 1 vaginal sample) (Supplementary Table S1). MfPV7, a member type within the α12 species, was the most common mucosal type found in 17 genital samples and 2 perianal samples (Supplementary Table S2). A number of α-PV types, including OTUαA12 (n = 19), OTUαA3 (n = 17), MfPV5 (n = 14), MfPV11 (n = 14) and MmPV6 (n = 11), commonly found in genital and/or perianal sites were absent in oral cavity. OTUαA21 (n = 11), OTUαA5 (n = 9) and OTUαA34 (n = 8) phylogenetically clustered into the γ-PV genus, however, were more common in oral samples, although they might infect perianal and/or genital sites occasionally (Supplementary Table S2).

Multiple-PV type infection in a same sample was common. 54.2% of PV-containing samples (77/142) were infected with at least two or more types, mostly common in genital swabs (63.1%, 41/65), followed by perianal (54.3%, 19/35), and oral (40.5%, 17/42) samples.

Eighteen macaque animals shared same PV type(s) between body sites. Identical types were found more commonly in perianal-genital pairs (63.1%, 41/65), followed by oral-perianal pairs (n = 6), and oral-genital pairs (n = 2). OTUαA12 and OTUαA28, two α-PV types, were mostly detected in 4 and 3 perianal-genital pairs, respectively. It is worth to note that infection of a same PV type crossing body sites could be from different sources but the close anatomic locations between genital and perianal areas likely facilitated the viral transmission from one body site to another more easily.

Detection of Novel Macaque PV Genomes
A BLAST search using partial L1 sequences against GenBank/NCBI and our PV database identified 25 potential novel α-PV types, 1 novel β-PV type, and 39 novel γ-PV types.

Table 1: Distribution of papillomavirus infection in a surveyed macaque population.

| Site         | Total No. (%) of samples tested | PV-positive No. (%) of samples within each genus | No. of unique PV types detected within each genus | Cumulative no. (%) of PV infection detected within each genus |
|--------------|---------------------------------|--------------------------------------------------|--------------------------------------------------|-----------------------------------------------------------|
|              | Total PV-positive                | α | β | γ | Total α | β | γ | Total α | β | γ | Total α | β | γ |
| Genital      | 117 (65.6)                      | 56 (86.2) | 1 (1.5) | 21 (32.3) | 52 | 32 | 1 | 19 | 198 | 159 (80.3) | 1 (0.5) | 38 (19.2) |
| Perianal     | 117 (29.9)                      | 18 (51.4) | 1 (2.9) | 18 (51.4) | 36 | 17 | 1 | 18 | 72 | 43 (59.7) | 1 (1.4) | 28 (38.9) |
| Oral         | 117 (35.9)                      | 3 (7.1) | 4 (9.5) | 38 (90.5) | 28 | 2 | 1 | 25 | 72 | 3 (4.2) | 4 (5.6) | 65 (90.3) |
| Total        | 351 (40.5)                      | 77 (54.2) | 6 (4.2) | 77 (54.2) | 72 | 32 | 1 | 39 | 342 | 205 (59.9) | 6 (1.8) | 131 (38.3) |

*Samples with PV infection from >1 genus were counted for each genus.*

are >300 (Alpha/papillomavirus tree) and >2000 (HPV variant trees of each type), indicating that all Bayesian chains were well sampled and have converged. The divergence times of non-human primate PVs partially sequenced by MY and FAP were estimated using the L1 substitution rate and the evolution models as mentioned above.
FIGURE 1 | Distribution of macaque PV genera and types in different body sites. (A) Pie charts showing the proportion of cumulative macaque PV infection between genera by gender and body sites. (B) Distribution of macaque PV infections in genital, anal and oral sites. The y-axis of the bar charts represents numbers of infections of each type. The phylogenetic tree on the bottom of the panel was inferred from previously reported macaque PV complete genomes and partial L1 sequences detected in this study. Animals with co-infections were counted more than once. Colors in the phylogenetic tree indicate different genera including Alpha-PV in red, Beta-PV in green, and Gamma-PV in blue.
Among them, 6 α-PV types (RhPVe, f, h, j, k, and m) and 6 γ-PV types (MfAA5, 7, 11, 20, 22, and 25) have been previously reported as partial MY and FAP sequences obtained from genital tissues (Chan et al., 1997) and skin swabs (Antonsson and Hansson, 2002, respectively).

The complete genomes of six macaque papillomaviruses (M. mulatta PV type 2, 3, 4, 5, 6, and 7) isolated from the surveyed animals have been previously characterized by the authors’ group. As shown in Supplementary Figure S1 of the predicted genomic structure, the viral genomes range from 7,292 bp to 7,935 bp in size, with five putative early genes (E6, E7, E1, E2, and E4), two late genes (L2 and L1) and an URR between the L1 and E6 genes. Although a small ORF (53 amino acids) between the E2 and L2 genes of MmPV2 and MmPV6 showed relatively high homology to the 3’-terminus of HPV6 E5a (33.9%), this gene warrants further characterization due to short sequence length. The L1 nucleotide sequence pairwise comparison and phylogenetic analysis indicated that these novel PV types shared less than 73% identities to the closest HPV relatives and well clustered into the genera α-PV (MmPV-2, -3, and -6) and γ-PV (MmPV-4, -5, and -7). Since all surveyed animals appeared healthy according to the body condition and the temperature measured by veterinarians, the pathogenic potential of these macaque PV types remains to be established.

Ancient Intra-Host Divergence of Primate PVs Prior to the Host Speciation

Following Fahrenholz's rule for strict virus-host codivergence (Hafner and Nadler, 1988), the evolutionary history of a pathogen should mirror that of its host, both in phylogenetic topology and divergence time. In order to estimate the split time between human and non-human primate PVs from their MRCAs, we applied a Bayesian statistical algorithm employing previously reported PV evolution rate and a combination of relaxed lognormal molecular clock and Yule population models. The divergence times were further calibrated by providing host fossil data in the trees, with assumption of codivergence histories between primate PVs and their hosts within specific phylogenetic clades. We observed a deep separation between Dyoomikron-PVs (represented by S. sciureus PV 1/2/3) and α-PVs split from a MRCA approximately 41 mya (95% highest posterior density, HPD, 37 – 44 mya) (Figure 4A and Table 2), coinciding with the time frame of host speciation between New World and Old World monkeys. Most of macaque PVs diverged from the MRCA with their closest HPV relatives around 14 – 33 mya, the time span roughly overlapping the speciation between macaque and human ancestors that occurred approximately 25 – 31 mya (Figure 4 and Supplementary Figures S4–6). Chimpanzee PVs (e.g., PpPV1) likely shared a MRCA with HPVs (e.g., HPV13) around 6 mya; this was also the time period that most of distinct HPV types emerged. The divergence times were further confirmed by the estimation including all α-, β- and γ-non-human primate PVs partially sequenced by MY and FAP (Supplementary Figure S7).

The tree topologies and divergence times support a model of ancient intra-host divergence of primate PVs in which multiple viral ancestors had split and adapted to specific host ecosystems (e.g., mucosal or cutaneous sites) within an ancestral host animal lineage (e.g., the MRCA of primate animals), prior to the speciation events of host species (Figure 5). Following periods of subsequent divergence and niche adaptation (e.g., different types of epithelial cells), distinct but phylogenetically related PV types were transmitted to similar host ecosystems by closely related host animals when the host speciation occurred. Intra-host viral divergence may occur at multiple time periods and/or...
host niche locations, resulting in the radiation observed in the extant primate PV tree where viruses sort by tissue tropism and/or biological property but not host species. The model of ancient viral divergence coupled to niche adaptation may explain in part the clustering of α9/11/12 (including HPV16, HPV34, and MfPV3) that evolved from a common ancestor colonizing the cervicovaginal regions of human and macaque ancestor and containing determinants associated with cancer.
FIGURE 4 | Divergence time estimation of primate papillomaviruses to their most recent common ancestors (MRCA). The times were estimated separately for each genus (A) Alpha-PVs, (B) Beta-PVs, and (C) Gamma-PVs. A Bayesian MCMC method was used to estimate the divergence time as described in section “Materials and Methods.” The branch lengths are proportional to the divergence times in million year ago (mya), with branches in orange referring to non-human primate papillomaviruses split from their MRCA with their closest HPV relatives. The bar on the nodes represent the 95% highest posterior density (HPD) interval for the divergence times (see details in Supplementary Figures S4–S6, respectively).

DISCUSSION

In this study, we surveyed a broad spectrum of α-, β-, and γ-PV infection from paired oral, perianal and genital swab samples collected from 117 non-captive healthy rhesus macaques using Next-Gen sequencing amplicon assays. Overall, high infection prevalence and genetic diversity of macaque PVs were detected. All identified macaque PV sequences clustered into the genera α-, β-, and γ-PV, with similar phylogenetic topologies as that of HPVs. Significant tissue tropism of macaque PVs were observed; the majority of PV types isolated from genital regions phylogenetically clustered into the genus α-PV, whereas the oral cavity harbored more γ-PVs. In consideration of low mutation rate of PV evolution and rare interspecies transmission, the close relationship between macaque PVs and HPVs with similar biological properties (e.g., α12 and α9 species associated with cancer) suggests a model of ancient divergence of primate PV ancestors adapting to specific host ecosystems prior to the host speciation. This phylogenetic topology was supported by a number of previously reported PV sequences from a wide range of apes and monkeys exclusively clustering into the genus α-, β-, and γ-PVs, in accordance to host anatomic sites where these viruses were originally isolated (Van Ranst et al., 1991; Chan et al., 1997; Antonsson and Hansson, 2002). Although a number of macaque PV types can be detected in multiple anatomic sites of the surveyed animals, making the actual tissue-tropism of primate PVs tentative, the self-transmission of viruses between the close anatomic locations (e.g., perianal and genital areas) should be a factor that could be ignored due to the poor sanitation of monkeys.

Using Bayesian molecular clock models, we estimated ancient intra-host divergence of primate PV ancestors within an ancestral host animal lineage occurred at least 41 million years ago, the era prior to the speciation between New World and Old World monkeys (Perez et al., 2013). Following subsequent divergence and niche adaptation, the speciation events of host animals (e.g., the split between macaques and humans) might isolate viruses with similar biological properties and/or niches into distinct but phylogenetically related clades (e.g., α9/11/12). This scenario may explain the close clustering of oncogenic human and non-human primate PVs (e.g., MfPV3 and HPV16) sharing a most recent common ancestor that contained the determinants for cervicovaginal colonization and cervical cancer. Interestingly, our unpublished data indicates that MfPV3 E6 can significantly degrade host p53 protein. It has been widely reported that high-risk HPV types (e.g., HPV16) can bind and degrade human p53 protein, a tumor suppressor gene targeted by viral E6 protein.
TABLE 2

| CALIBRATION MODEL | TREE | AICM Tree Prior | MRCA (Million Years Ago, mya) | log Alpha vs. HPD | 95% interval | 95% interval | AICM value | AICM Difference |
|-------------------|------|----------------|-------------------------------|------------------|-------------|-------------|------------|----------------|
|                   | 3 Call | Relaxed Bayesian | 712324 | 164 | -0.559065 | 40.7 | 29.7 | 21.9 | 22.0 | 25.9 | [19.4, 25.5] | [17.4, 17.7] | [14.3, 14.6] | [9.6, 9.8] |
|                   | 3 Call | Relaxed Constant | 712323 | 465 | 44.3 | 18.7 | 18.8 | 22.2 | 25.3 | [19.4, 21.7] | [17.4, 24.9] | [14.6, 17.1] | [9.6, 9.8] |
|                   | 3 Call | Strict Bayesian | 717187 | 1055 | -0.559464 | 40.6 | 29.8 | 24.4 | 21.9 | 21.8 | [19.6, 24.9] | [17.4, 20.4] | [14.6, 17.1] | [9.6, 9.8] |
|                   | 3 Call | Strict Constant | 717176 | 1055 | 44.3 | 21.7 | 21.8 | 24.4 | 20.9 | [19.6, 24.8] | [17.4, 20.4] | [14.6, 17.1] | [9.6, 9.8] |

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

Chi-, Lambda- and Tau-PVs, although these viruses were mainly

September 2019 | Volume 10 | Article 2093

Diversification of PV genomes independent from host speciation can occur in non-primate animals harboring many PV types belonging to different genera, such as bovine, canine and feline PVs. At least 27 distinct BPV types have been characterized and were phylogenetically assigned into five genera including Delta- (BPV1, 2, 13 and 14), Xi- (BPV3, 4, 6, 9, 10, 11, 12, 15, 17, 20, 24, 26, and 04AC14), Episilon- (BPV5, 8, and 25), Dyoxi- (BPV7, 19, 21, and ujs21015), and Dyokappa-PV (BPV16, 18 and NY8385) (Rector and Van Ranst, 2013; Van Doorslaer et al., 2018). Bovine PVs have been reported to be associated with mucocutaneous papillomas and neoplasia of the bladder and the upper alimentary tract (Gil da Costa et al., 2017). Interestingly, except for the genus Dyoxi-PV, the rest four genera also contained the majority of PV types isolated from artiodactylous animals, such as Ovis aries PVs, Camelus dromedaries PVs, and Cervus elaphus PVs, suggesting an ancient intra-host viral divergence history occurred in the ancestors of artiodactylous PVs. Following virus-host coinfection within specific host niches, artiodactylous PVs infecting a single host species might form distinct clade that was phylogenetically related to viruses from other closely related host species. We found the similar evolutionary pattern in canine and feline PVs that distributed into the genera Chi-, Lambda- and Tau-PV, although these viruses were mainly

(Schiffman et al., 2016). The data indicates that certain types of macaque PVs may carry carcinogenic property similar as high-risk HPV types do; the close phylogenetic relationship between α12 and α9 species supports evolutionary adaptation of biological functions between human and non-human primate PVs from their common ancestors.

Virus-host codivergence was generally responsible for the large evolutionary distances between PVs belonging to different order and/or class (e.g., mammal vs. feline). However, this term does not explain the distribution of primate PVs infecting a single host species into different phylogenetic clades. Ancient intra-host viral divergence undergoing slow genomic mutations in respond to selection pressure by host immune responses and viral genetic drift may have properties beneficial for virus strategies, such as improved infectivity, regulation and pathogenesis within specific host niches. For example, all α-PVs but not β-/γ-PVs contain a short region in length of 300 – 500 bp between E2 and L2 genes, including an E5 ORF sharing high homology between genital high-risk HPVs and MiPVs (Bravo and Alonso, 2004; Chen et al., 2009). The E5 protein has been reported to be important in host-cell transformation (DiMaio and Mattoo, 2001). Niche adaptation is a slow and long-term progress, restricted by ecological and host environment as well as molecular interplaying between virus and host regulatory proteins (Shadan and Villarreal, 1993; Bravo and Felez-Sanchez, 2015). Subsequent speciation events of the host animals were able to isolate virus ancestors within niche-specific clades into different populations, with or without phenotypic differences. The bottleneck of virus transmission between human and non-human primate hosts formed all observed macaque PVs, for example, into distinct clades (e.g., the species α12) excluding any of HPV types, and vice versa, with mirrored phylogenetic topologies and divergence times between hosts and parasites.
isolated from the oral cavity and their pathogenesis warrants further investigation.

It is unclear whether viral transmission occurred between human and non-human primate ancestors. Although there is no report of any HPV DNA being detected in an animal or of any animal PV DNA in a human, the extreme species restriction is probably exclusively operative in primate PVs. Given close cutaneous or mucosal contacts between hominid and macaque ancestors in the early stage of host speciation, host interbreeding and switching via horizontal transmission may be responsible in part for a recent split of MmPV5 and CgPV2 from their closest HPV relatives, as estimated around 13 – 14 mya.

Non-human primates share various behavioral, physiological and genetic similarities with humans, making monkey PVs as ideal animal models to understand the origin and evolution of HPV carcinogenicity. A number of macaque PV types within the species \( \alpha_{12} \) isolated from mucosal lesions of rhesus and cynomolgus macaques have been associated with cervical neoplasia (Wood et al., 2007). The evolutionary relatedness between \( \alpha_{12} \) and \( \alpha_{9} \) species highlights macaque PVs serving as a unique and highly relevant preclinical model that minimizes limitations of mucosal tropism, viral transcription and genomic factors for the study of HPV persistence and oncogenesis (Campo, 2002; Hu et al., 2017).

One of the limitations of this study is the short of cytological and/or histological information to define the potential pathogenicity of macaque PVs in the surveyed macaque population. The most predominant genital macaque PV was MfPV7 found in 19 animals (16%), a member type of the species \( \alpha_{12} \) containing MfPV3 associated with cervical cancer. However, the oncogenic potential of MfPV7 and/or other closely related types within the species \( \alpha_{12} \) is not clearly known. Although the L1-based amplicon sequencing assays were able to detect a broad spectrum of primate PVs, the implementation of whole genome metagenomic sequencing may provide a deep genomic coverage to further explore viral mutation and diversity (Pastrana et al., 2018; Tirosh et al., 2018). We only sampled a single macaque species inhabiting a protected nature reserve. It is possible that an increased diversification of macaque PV types, particular the \( \beta \)-PVs could be observed in populations from different geographic locations and macaque species. To our knowledge, however, this is the most comprehensive study of primate PV genomic heterogeneity and evolution by including all up-to-date non-human primate PV complete genomes and partial sequences, which strengthens the model of ancient intrahost divergence of primate PVs.

CONCLUSION

In conclusion, this study establishes a comprehensive theoretical framework on the evolution of primate PVs, suggesting an ancient intrahost viral divergence model upon niche adaptation followed by virus-host codivergence. Our model provides a genetic foundation to better understand the origin and evolution of HPV pathogenesis and carcinogenicity associated with cervical cancer, and the underlying contribution of host niche adaptation to viral fitness. The oncogenic outcomes of human and non-human primate PVs imply an evolutionary convergence of virus-host interaction involving complex interaction between host and pathogens in association with different genotypes, phylotypes and phenotypes.

DATA AVAILABILITY

Accession numbers for the sequences determined in this study are available in GenBank under the accession numbers MG837557, MG837558, MG837559, MH745747, MH745748, and MH745749.
ETHICS STATEMENT

The animal use protocol has been approved by the Agricultural and Fishery Department and presented to the Animal Welfare Advisory Group (AWAG). All experiments were performed in accordance with the relevant guidelines and regulations.

AUTHOR CONTRIBUTIONS

All authors were involved in the conception and design of the study, and approved the final manuscript. ZC, PC, and RB contributed to the study design and supervision. TL, PW, and WH processed the samples and performed the sequencing. ZC and TL contributed to the data analysis and wrote the original draft of the manuscript. PC and RB reviewed and edited the manuscript.

FUNDING

This work was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 24100716) to ZC. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

ACKNOWLEDGMENTS

We thank the Agriculture, Fisheries and Conservation Department of the Government of the Hong Kong Special Administrative Region for providing access to the macaques, and the Ocean Park Conservation Foundation Hong Kong for assistance with sample collection. We also thank Mr. Samuel Tong and Jeffrey Xu for the assistance with sample transportation and processing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.02093/full#supplementary-material

REFERENCES

Agalliu, I., Gapsur, S., Chen, Z., Wang, T., Anderson, R. L., Teras, L., et al. (2016). Associations of oral alpha-, beta-, and gamma-human papillomavirus types with risk of incident head and neck cancer. JAMA Otolaryngol. 2, 599–606. doi: 10.1001/jamaoncol.2015.5304

sequence alignment of 6 open reading frames (ORF6) (E6-E7-E1-E2-L2-L1) of 145 papillomavirus types representing 136 species and unique host species (see PV list in Supplementary Table S3, column of “Selected type”). The number on the nodes indicate maximum likelihood bootstrap percentages. The main clades containing the majority of primate papillomavirus species are highlighted in gray.

FIGURE S1 | Genomic structure of six Macaca mulatta papillomavirus complete genomes.

FIGURE S2 | A maximum likelihood (ML) phylogenetic tree of papillomaviruses. The tree was constructed using RAxML, based on the concatenated nucleotide sequence alignment of 6 open reading frames (ORF6) (E6-E7-E1-E2-L2-L1) of 145 papillomavirus types representing 136 species and unique host species (see PV list in Supplementary Table S3, column of “Selected type”). The number on the nodes indicate maximum likelihood bootstrap percentages. The main clades containing the majority of primate papillomavirus species are highlighted in gray.

FIGURE S3 | Phylogenetic incongruence of papillomaviruses. ML phylogenetic trees were constructed using the concatenated nucleotide sequence alignment of early genes (E1-E2) and late genes (L2-L1) of 145 papillomavirus types representing 136 species and unique host species (see PV list in Supplementary Table S3). The branches in orange represent non-primate papillomaviruses. Non-primate animal papillomaviruses are collapsed in gray. Numbers on the branches indicate support indices of maximum likelihood bootstrap percentages using RAxML, and Bayesian credibility value percentage using MrBayes. 100% agreement is denoted by an asterisk (*).

FIGURE S4 | Divergence time estimation of Alphapapillomaviruses and Dyoomikronpapillomaviruses to their most recent common ancestors (MRCAs). A Bayesian MCMC method was used to estimate the divergence time of Alphapapillomaviruses and Dyoomikronpapillomaviruses from their most recent common ancestors. The branch lengths are proportional to the divergence times. The branches in orange refer to non-human primate papillomaviruses. The number on the nodes are the mean estimated divergence time in million year ago (mya). The bar on the nodes represent the 95% highest posterior density (HPD) interval for the divergence times.

FIGURE S5 | Divergence time estimation of Betapapillomaviruses to their most recent common ancestors. A Bayesian MCMC method was used to estimate the divergence time of Betapapillomaviruses from their most recent common ancestors. The branch lengths are proportional to the divergence times. The branches in orange refer to non-human primate papillomaviruses. The number on the nodes are the mean estimated divergence time in mya. The bar on the nodes represent the 95% HPD interval for the divergence times.

FIGURE S6 | Divergence time estimation of Gammapapillomaviruses to their most recent common ancestors. A Bayesian MCMC method was used to estimate the divergence time of Gammapapillomaviruses from their most recent common ancestors. The branch lengths are proportional to the divergence times. The branches in orange refer to non-human primate papillomaviruses. The number on the nodes are the mean estimated divergence time in mya. The bar on the nodes represent the 95% HPD interval for the divergence times.

FIGURE S7 | Divergence time estimation of Alphapapillomaviruses and Dyoomikronpapillomaviruses (A), Betapapillomaviruses (B), and Gammapapillomaviruses (C) inferred from partial L1 sequences by MY and FAP. A Bayesian MCMC method was used to estimate the divergence time of non-human primate PVs and HPVs from their most recent common ancestors. The branch lengths are proportional to the divergence times. The branches in orange refer to non-human primate papillomaviruses. The number on the nodes are the mean estimated divergence time in mya. The bar on the nodes represent the 95% HPD interval for the divergence times.

TABLE S1 | Papillomavirus infection in a surveyed macaque individuals.

TABLE S2 | Macaque papillomavirus types infecting in a surveyed macaque population.

TABLE S3 | List of papillomavirus complete genomes used in this study.

TABLE S4 | List of partial L1 gene of papillomavirus types analyzed in this study.

Antonsson, A., and Hansson, B. G. (2002). Healthy skin of many animal species harbors papillomaviruses which are closely related to their human counterparts. J. Virol. 76, 12537–12542. doi: 10.1128/jvi.76.24.12537-12542.2002

Bael, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., and Aleskeyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating
Chen et al. Evolution of Primate Papillomaviruses

Sharp, P. M., and Simmonds, P. (2011). Evaluating the evidence for virus/host co-evolution. *Curr. Opin. Virol.* 1, 436–441. doi: 10.1016/j.coviro.2011.10.018

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Tirosh, O., Conlan, S., Deming, C., Lee-Lin, S. Q., Huang, X., Program, N. C. S., et al. (2018). Expanded skin virome in DOCK8-deficient patients. *Nat. Med.* 24, 1815–1821. doi: 10.1038/s41591-018-0211-7

Van Doorslaer, K., Chen, Z., Bernard, H. U., Chan, P. K. S., Desalle, R., Dillner, J., et al. (2018). ICTV virus taxonomy profile: Papillomaviridae. *J. Gen. Virol.* 99, 989–990. doi: 10.1099/jgv.0.001105

Van Doorslaer, K., Ruoppolo, V., Schmidt, A., LesCroel, A., Jongsoomjit, D., Elrod, M., et al. (2017). Unique genome organization of non-mammalian papillomaviruses provides insights into the evolution of viral early proteins. *Virus Evol.* 3:vex027. doi: 10.1093/viruses/vex027

Van Ranst, M., Fuse, A., Sobis, H., De Meurichty, W., Syrjanen, S. M., Billiau, A., et al. (1991). A papillomavirus related to HPV type 13 in oral focal epithelial hyperplasia in the pygmy chimpanzee. *J. Oral Pathol. Med.* 20, 325–331. doi: 10.1111/j.1600-0714.1991.tb00939.x

Wong, M. C. S., Vlantis, A. C., Liang, M., Wong, P. Y., Ho, W. C. S., Boon, S. S., et al. (2018). Prevalence and epidemiologic profile of oral infection with alpha, beta, and gamma Papillomaviruses in an Asian Chinese Population. *J. Infect. Dis.* 218, 388–397. doi: 10.1093/infdis/jiy160

Wood, C. E., Chen, Z., Cline, J. M., Miller, B. E., and Burk, R. D. (2007). Characterization and experimental transmission of an oncogenic papillomavirus in female macaques. *J. Virol.* 81, 6339–6345. doi: 10.1128/jvi.00233–07

Woolhouse, M. E., Webster, J. P., Domingo, E., Charlesworth, B., and Levin, B. R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nat. Genet.* 32, 569–577. doi: 10.1038/ng1202-569

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

*Copyright © 2019 Chen, Long, Wong, Ho, Burk and Chan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*