Virology 445 (2013) 11–20

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Evolution of the *Papillomaviridae*

Koenraad Van Doorslaer

DNA Tumor Virus Section, Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 209892, USA

ARTICLE INFO

Article history:
Received 21 March 2013
Returned to author for revisions
2 April 2013
Accepted 9 May 2013
Available online 14 June 2013

Keywords:
Papillomaviridae
Co-evolution
Evolution
Niche adaptation
Recombination
Codon usage
Mutation rate

ABSTRACT

Viruses belonging to the *Papillomaviridae* family have been isolated from a variety of mammals, birds and non-avian reptiles. It is likely that most, if not all, amniotes carry a broad array of viral types. To date, the complete genomic sequence of more than 240 distinct viral types has been characterized at the nucleotide level. The analysis of this sequence information has begun to shed light on the evolutionary history of this important virus family. The available data suggests that many different evolutionary mechanisms have influenced the papillomavirus phylogenetic tree. Increasing evidence supports that the ancestral papillomavirus initially specialized to infect different ecological niches on the host. This episode of niche sorting was followed by extensive episodes of co-speciation with the host. This review attempts to summarize our current understanding of the papillomavirus evolution.

Published by Elsevier Inc.

Introduction

Papillomaviruses have been discovered in a wide array of vertebrates (Rector and Van Ranst, in press, and references therein). Remarkably, in those hosts that have been extensively studied (e.g. humans, cattle, dogs, etc.), an extensive repertoire of highly species-specific viruses has been described. With over 240 distinct types classified in 37 genera, papillomaviruses may be the biggest and most successful family of vertebrate viruses (Bernard et al., 2010; de Villiers et al., 2004; Van Doorslaer et al., 2013). It is clear that papillomaviruses have been, and continue to be, an astonishing evolutionary success.

The double stranded papillomavirus genome ranges from 6953 bp [Chelonia mydas papillomavirus type 1 (CmPV1)] to 8607 bp [Canine papillomavirus type 1 (CPV1)] in length. Due to their clinical importance, the relatively small genomic size and recent advances in cloning and sequencing methods the complete genomic sequences of most identified viral types is available to researchers. This treasure-trove of complete genomic sequence data in combination with clinical and biochemical data has made papillomaviruses an ideal model system to understand how evolution affects different aspects of the viral lifecycle. It has been shown that papillomaviruses have a slow evolutionary rate, suggesting that genetic drift is mainly responsible for viral diversity. Indeed, given that papillomaviruses infect their hosts for very long periods of time, it remains an open question what additional types of evolutionary pressure (if any) should be expected to shape the viral genome.

This review presents an account of our current understanding of the papillomavirus evolutionary history. The more we learn about the evolution of papillomaviruses, the more "nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973) rings true.

Papillomaviruses co-speciated with their hosts

Historically, viral evolution has mainly been considered from a predator–prey perspective. Under this model, viral fitness (and thus its evolutionary success) is measured by the viral capacity to cause disease in its host (Shadan and Villarreal, 1995). However, papillomaviruses (and other small DNA viruses) cause benign, mostly unapparent, persistent infections in their hosts. In addition, papillomaviruses are highly host-restricted, and cause abortive infections in non-host species. In fact, the only exceptions to strict species specificity were described in mammalian hosts known to hybridize (Chen et al., 2009; Gottschling et al., 2011a), thereby challenging the hosts' species definition. The observation that papillomaviruses cause benign infections unable to cross the hosts' species-barrier has led to the hypothesis of "host-linked evolution" (Bernard, 1994; Sundberg et al., 1997).

The traditional (orthogenetic) definition of co-evolution states that parasites of closely related host species should be closely related themselves and cluster together in the parasite phylogenetic tree. Furthermore, dates associated with parasite

E-mail address: Koenraad.vandoorslaer@nih.gov

0042-6822/$ - see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.virol.2013.05.012

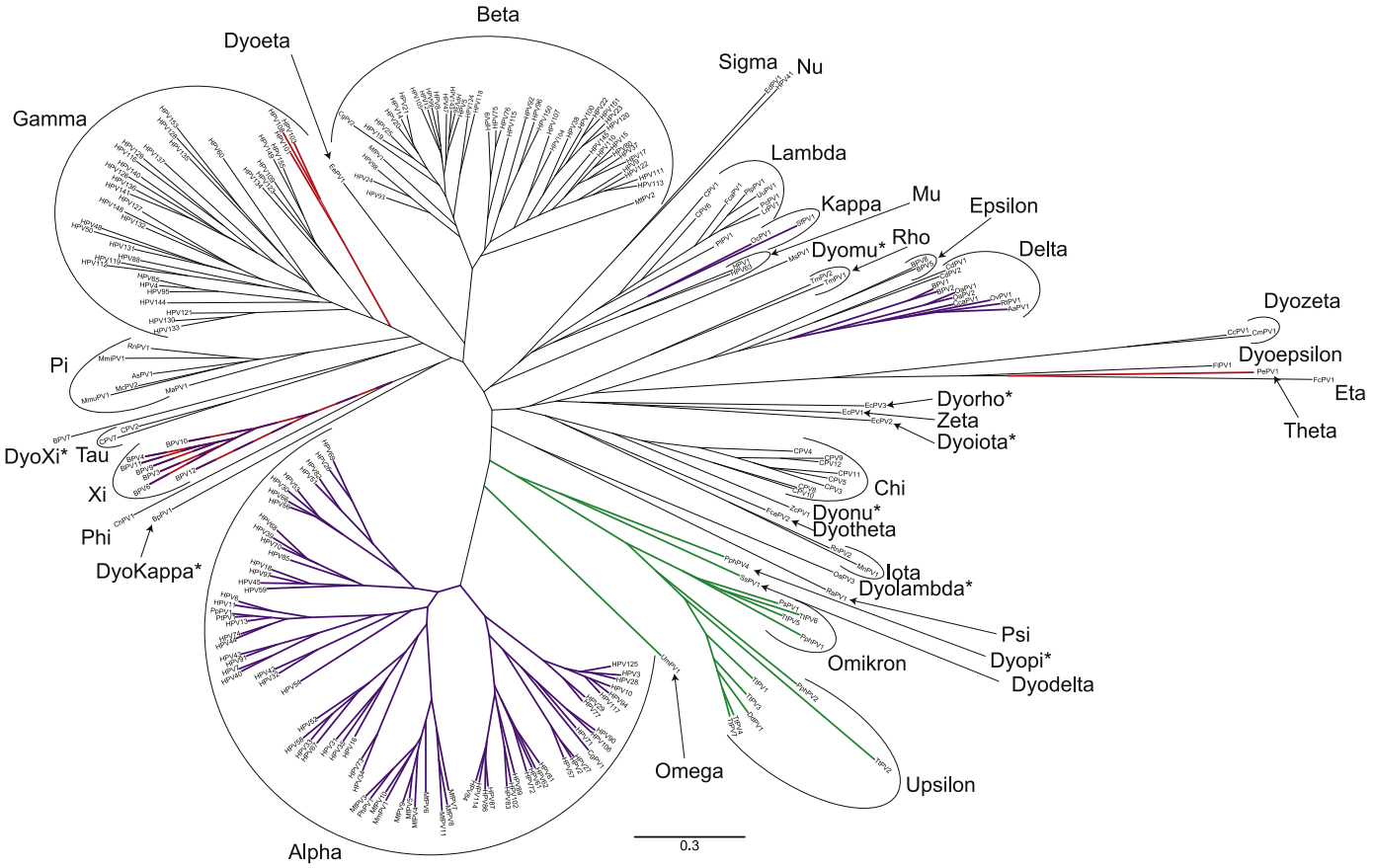

Fig. 1. Papillomavirus phylogenetic tree. The DNA sequence coding for E1, E2, L1 and L2 for all 241 papillomaviruses currently on PaVE were downloaded and aligned. A partitioned gene alignment was used as the base for a maximum likelihood reconstruction of the phylogenetic tree. The different papillomavirus genera are named according to Bernard et al. (2010) and de Villiers et al. (2004). Genera marked with an asterisk have been proposed to the ICTV, and are awaiting official recognition (http://talk.ictvonline.org/files/proposals/taxonomy_proposals_vertebrate1/m/vert01/4244.aspx). The tree is color-coded according to presence/absence of the "adaptive proteins". Red clades lack an E6 ORF. The viruses highlighted in green do not code for an E7 protein. The purple clades code for a hydrophobic E5 protein. The *Xipapillomaviruses* lack an E6 (red), but contain an E5 (purple).

divergence should coincide with the host-species divergence (Fahrenholz's rule; Hafner and Nadler, 1988). Therefore, any incongruence between both trees should be considered as evidence that parasite and host did not co-evolve (Brooks, 2003; Page, 2002).

With an increase in the number of papillomavirus sequences (and their associated hosts), it became clear that papillomaviruses and their hosts did not follow an identical evolutionary path (Bravo and Alonso, 2007; Chan et al., 1992, 1995, 1997a, 1997b; Garcia-Vallve et al., 2005; Gottschling et al., 2007a, 2007b; Narechania et al., 2005a; Rector et al., 2008). Several violations of strict co-evolution can be observed in the phylogenetic tree in Fig. 1. For example, human papillomaviruses can be found in five different genera (Alpha, Beta, Gamma, Mu and Nu) dispersed throughout the phylogenetic tree. Also, strict co-evolution would place the non-human primate papillomaviruses basal to human papillomaviruses, not intermingled as is observed.

Evolutionary events such as cross-species infection, recombination and virus duplication (e.g. following ecological niche adaptation) have been suggested to explain the observed conflicts (Angulo and Carvajal-Rodriguez, 2007; Carvajal-Rodriguez, 2008; Gottschling et al., 2011b; Halpern, 1995; Shah et al., 2010; Varsani et al., 2006). The impact of viral recombination will be discussed in more detail later, but while the influence of recombination has been limited, recombination likely played a role at specific moments throughout viral evolution.

Because of the absence of cross-species infections, it is unlikely that horizontal gene transfer played any role in the evolution of the *Papillomaviridae*. In fact, a study specifically looking at the influence of horizontal gene transfer identified only a single potential cross-species transmission event. This event involved ancestors of a porcupine (EdPV1) and human (HPV41) papillomavirus (Shah et al., 2010). These two viruses are the only members of a divergent genus (*Nupapillomavirus*); it will be of interest to see how the inclusion of more viruses in this genus will affect the conclusion of cross-species infection.

A more recent version of the co-evolution theory was initially proposed in the early 1960s (reviewed in Brooks and Ferrao, 2005). This updated theory states that the evolution of parasites follows the evolution of host resources, not the evolution of the host species per-se. The shape of the papillomavirus phylogenetic tree could potentially be explained using this interpretation of co-evolution. Under this model, specific events in the evolution of hosts (e.g. presence/absence of fur, evolution of sweat glands, etc.) created new ecological niches for papillomaviruses to adopt (Bernard et al., 1994). Therefore, the data suggests a model in which a generalist ancestral papillomavirus diverged into four or five increasingly specialized viruses (reflected in the 4–5 major clades of the phylogenetic tree) (Bravo and Alonso, 2007). Following these niche adaptation events, the virus evolved alongside its hosts. Throughout the co-evolutionary process, the availability of new niches would in turn drive viral radiation, followed by further co-speciation.

In conclusion, the papillomavirus phylogenetic tree cannot be explained solely by co-evolution. However, initial niche sorting followed by virus–host linked speciation was a key determinant of the papillomavirus evolutionary history (Gottschling et al., 2011b; Shah et al., 2010).

Table 1
Papillomaviruses gained and lost "adaptive proteins" several times throughout evolution.

Genus (species)	E6	E7	E5
Deltapapillomavirus (1–5)	Yes	Yes	Yes
Gammapapillomavirus (6)	No	Yes	No
Upsilonpapillomavirus (all)	Yes	No	No
Omikronpapillomavirus (all)	Yes	No	No
Omegapapillomavirus (all)	Yes	No	No
Dyopipapillomavirus (all)	Yes	No	No
Dyodeltapapillomavirus (all)	Yes	No	No
Xipapillomavirus (all)	No	Yes	Yes*
Alphapapillomavirus (5–7, 9 and 11)	Yes	Yes	E5α
Alphapapillomavirus (2–4 and 14)	Yes	Yes	E5β
Alphapapillomavirus (10)	Yes	Yes	E5γ and E5δ
Alphapapillomavirus (8)	Yes	Yes	E5δ
Alphapapillomavirus (12)	Yes	Yes	Yes[&]
Kappapapillomavirus	Yes	Yes	Yes

Not all extant papillomaviruses encode an E1, E2 (and E4), E6, E7, L2 and L1 ORF. Viral genera (species indicated by number) that gained an E6, E7 or lost an E5 ORF are listed. The *Alphapapillomavirus* E5 ORFs are named according to Bravo and Alonso (2004). Please note that the species nomenclature of the *Alphapapillomaviridae* was updated in 2010 (Bernard et al., 2010), explaining some discrepancies between Table 1 and the paper by Bravo and Alonso (2004). The E5 proteins found in viruses belonging to the Alphapapillomavirus species 12 (indicated by [&]) were not discussed in the paper by Bravo and Alonso (2004), and were therefore not specifically named. The E8 protein encoded by the members of the *Xipapillomaviruses* was renamed E5 (indicated by asterisk).

Gains (and losses) of "adaptive proteins" drive papillomavirus evolution

All known papillomaviruses encode for at least five proteins (E1, E2 (and E4), L1 and L2). E1 and E2 are key modulators of replication and transcription, while the L1 and L2 structural proteins make up the viral capsid. The E4 protein is embedded within the E2 protein and is expressed as a very abundantly spliced transcript. In addition to this core set of proteins, all papillomaviruses contain an untranslated long control region (LCR) located between the L1 and E6 ORFs. The LCR contains elements required for transcription and replication (Garcia-Vallve et al., 2005). Most viruses also encode proteins (E5, E6 and E7) involved in modulating cellular growth and immune responses (Klingelhutz and Roman, 2012). In the cancer-associated viruses, these proteins have been shown to be oncogenic. The phylogenetic tree in Fig. 1 illustrates that certain evolutionary groups are "apomorphic" for the presence/absence of these non-core proteins (green clades lack an E7 protein; red clades do not encode an E6. The *Alpha-*, *Delta-* and *Kappapapillomaviruses* contain an E5 ORF; summarized in Table 1). Theoretically, a putative virus consisting of the core set of features (LCR, E1, E2, L2 and L1) should be able to fulfill the basic requirements (replication, transcription and viral packaging) of a viral infection. It is therefore tempting to hypothesize that the ancestral papillomavirus did not contain any "adaptive proteins" (E5, E6 and/or E7). However, all known viruses contain at least one of these proteins (Table 1 and Fig. 2).

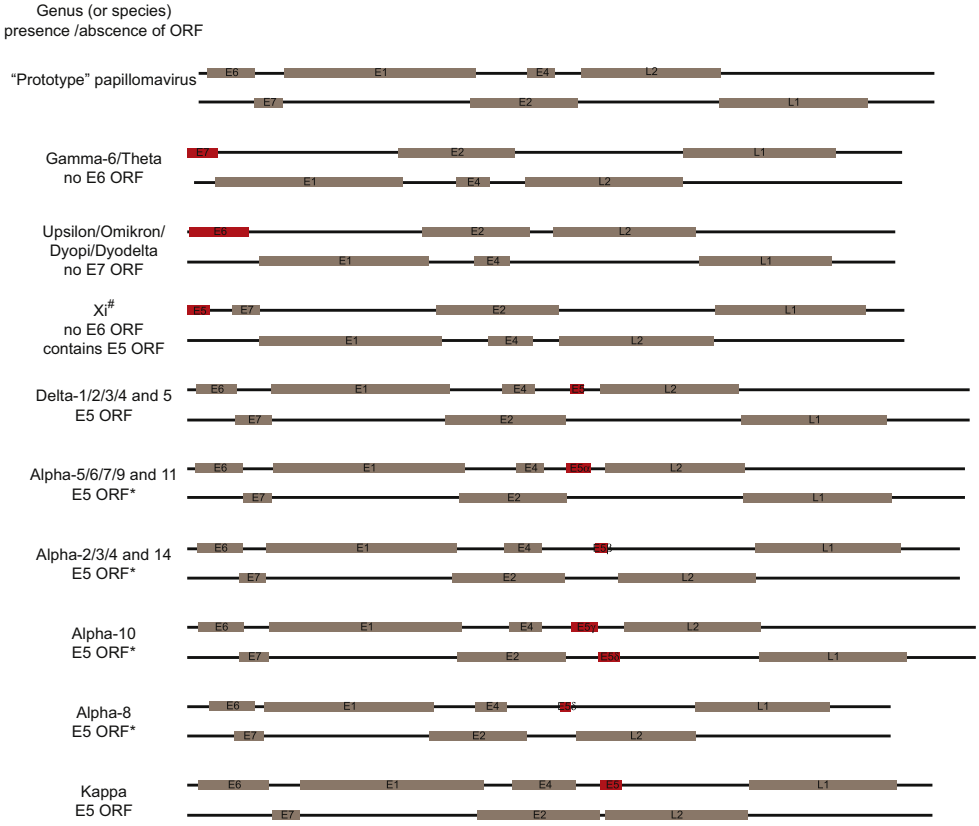

Fig. 2. Gains and losses of "adaptive proteins" throughout papillomavirus evolution. Graphical representation of different viral genomes illustrating the loss and/or gain of viral proteins. Each ORF is represented by a rectangle. Viral genera (or species) are indicated at the left, and a representative viral type is used to illustrate genomic features. The red boxes indicate viral ORFs different from the prototypical PV genome. The E5 proteins found in members of the *Alphapapillomaviruses* (indicated by "*") are labeled according to Bravo and Alonso (2004). The E5 proteins found in viruses in the genus *Xipapillomaviruses* (indicated by "#") were previously named E8 (O'Brien et al., 2001).

While it is hard to speculate on the evolutionary origin of these individual building blocks, it is likely that at least some of these components were already available for incorporation into (papilloma-) viral genomes. This idea is supported by the relatively high sequence similarity between the helicase motifs of the papillomavirus E1 protein, the polyomavirus large T-antigen (Clertant and Seif, 1984), the parvovirus NS1 protein (Astell et al., 1987), and a virus-like element found in certain flatworms (Rebrikov et al., 2002). While it is impossible to exclude convergent evolution, it is tempting to speculate on a shared evolutionary history of these viral proteins (Bernard, 2013).

The mammalian E6 proteins contain two highly conserved 70-residue zinc-binding repeats (Zanier et al., 2013). Interestingly, the avian papillomavirus E6 protein consist of a single zinc-binding domain that is closely related to the C-terminal domain of the mammalian E6 protein. Phylogenetic evidence points towards the existence of a comparable single-domain E6 protein in the proto-papillomavirus (Van Doorslaer et al., 2009). The mammalian E6 N-terminal domains may have emerged by duplication and subsequently diverged from the original ancestral domain (Van Doorslaer et al., 2009). The mammalian E7 proteins contain an unfolded N-terminal region followed by a zinc-binding motif (Liu et al., 2006; Ohlenschlager et al., 2006). Although the primary sequence of the E7 zinc-binding domain is highly reminiscent of the E6 domain (Cole and Danos, 1987), it has a drastically different fold. While it is possible that the different E6 and E7 proteins are the result of convergent evolution, the more parsimonious explanation suggests that an "E6/E7-like" protein became part of the ancestral papillomavirus genome prior to the initial niche adaptation and divergence at the root of the tree (see above). While the order of events has not been clarified, it is likely that the extant E6 and E7 proteins arose through a series of duplication events after which each protein diverged and adopted new functions (Fig. 3). The low overall sequence similarity between E6 or E7 protein from different viruses supports the notion that these proteins have evolved to fulfill a diverse set of functions depending on the needs of a particular viral type. Indeed, recent proteomic studies confirm that the set of cellular proteins targeted by the E6 and/or E7 proteins are unique to a specific viral type (Rozenblatt-Rosen et al., 2012; White and Howley, 2013; White et al., 2012a, 2012b). Remarkably, these adaptations may manifest at the structural level. The recently solved human (HPV16) and bovine (BPV1) E6 structures show that both proteins evolved drastically different ways to interact with their cellular targets (Zanier et al., 2013). This raises the possibility that the functions performed by these proteins are not strictly essential to the viral lifecycle, but that changes to these "adaptive proteins" allowed viruses to adjust to new eco-systems on their hosts.

Given that the E6/E7 protein was acquired early in the evolution of the *Papillomaviridae*, the absence of these proteins in certain viruses (Fig. 1) implies a loss of the ORF. Interestingly, the loss of E6 appears to have occurred repeatedly throughout viral evolution (Fig. 1, red clades). On the other hand, based on the current sequence information, loss of E7 has only befallen related

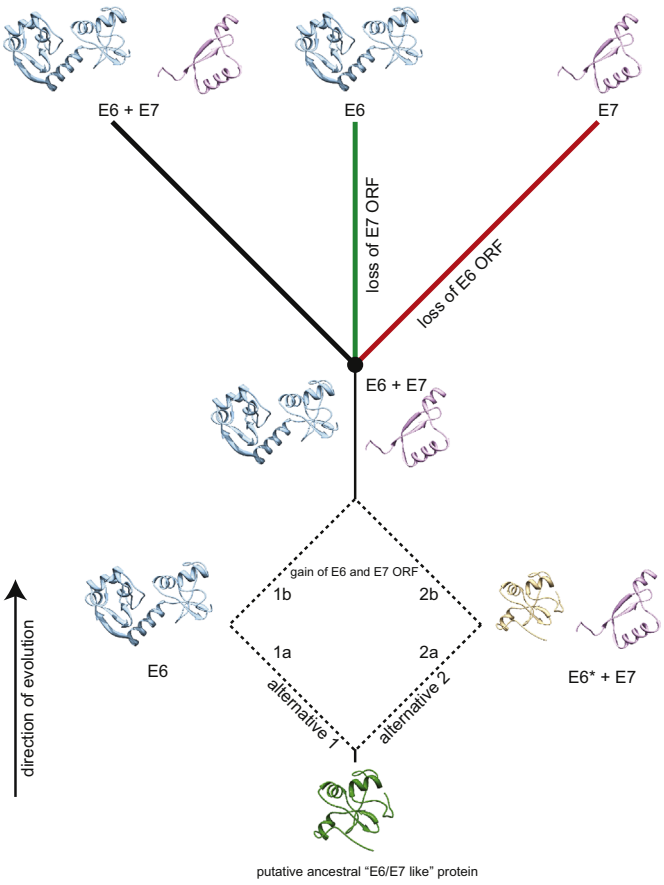

Fig. 3. Evolution of the E6 and E7 proteins. The proto-papillomavirus may have contained a hypothetical "E6/E7 like" protein. Phylogenetic evidence suggests that this ancestral protein was similar to the single-domain E6 protein present in extant avian papillomaviruses (FlPV1 and FcPV1) (Van Doorslaer et al., 2009). The E6 C-terminal domains present in extant mammalian PVs are phylogenetically related to these single-domain avian E6. Like the E6 proteins, the E7 protein contains a highly conserved Zinc binding motif. The presence of a conserved functional motif suggests common ancestry the E6 and E7 proteins (Cole and Danos, 1987). In alternative pathway 1, the ancestral "E6/E7 like" protein duplicated, creating the two-domain E6 protein. The N-terminal domain subsequently diverged from the original ancestral domain (path 1a). This initial duplication was followed by an additional duplication that created the ancestral E7 protein. Further evolutionary processes shaped the E7 protein (path 1b). In alternative pathway 2, the duplication event giving rise to the E7 protein (path 2a) preceded the domain duplication that created the two domain E6 (path 2b). Throughout evolution some viruses lost the E7 or the E6 protein (green and red branch respectively).

Table 2
Summary of papillomavirus codon sites under positive selection.

ORF	Codon position (amino acid)	Reference
HPV16		
E6	17 Arg	DeFilippis et al. (2002), Chen et al. (2005) and Carvajal-Rodriguez (2008)
	21 Gln	DeFilippis et al. (2002), Chen et al. (2005) and Carvajal-Rodriguez (2008)
	34 Ile	DeFilippis et al. (2002)
	90 Leu	DeFilippis et al. (2002), Chen et al. (2005) and Carvajal-Rodriguez (2008)
E5	48 Leu*	Chen et al. (2005)
	65 Ile*	Chen et al. (2005)
L2	378 Ser	Carvajal-Rodriguez (2008)
HPV18		
E5	72 Leu	Chen et al. (2009)
HPV45		
E6	21 Leu	Chen et al. (2009)
L1	357 Ser*	Chen et al. (2009)

Codon sites under diversifying selection (dN/dS > 1) have been described in HPV16, HPV18 and HPV45. The codon position and encoded amino acid are indicated. In some cases (indicated with asterisk), the codon positions were renumbered in order to be in agreement with the data in the PaVE database (Van Doorslaer et al., 2013).

viruses at the base of the genus *Alphapapillomavirus* (Fig. 1, green clades).

While an early evolutionary event added an "E6/E7 like protein" to the papillomavirus repertoire, several lines of evidence point towards several entry events of the viral E5 ORF (Garcia-Vallve et al., 2005). First, only members of the *Alphapapillomavirus*, *Deltapapillomavirus* and possibly *Kappapapillomavirus* genera contain an ORF coding for a hydrophobic protein in the traditional location (between the E2 and L2 ORFs; Fig. 2) (Garcia-Vallve et al., 2005). Certain members of the *Xipapillomavirus* genus lack an E6 gene and instead have a short hydrophobic ORF in the same genomic location. This ORF was originally termed E8 but shares characteristics of the E5 ORFs of the *Deltapapillomavirus* genus (O'Brien et al., 2001). Second, besides being hydrophobic in nature, these proteins do not share any chemical features and do not appear to be homologous. For example, all members of genus *Alphapapillomavirus* encode an E5 ORF, however these proteins can be divided into four separate groups with diverse chemical and evolutionary characteristics (Bravo and Alonso, 2004). Consistent with its recent acquirement by the genome, the E5 protein is one of few papillomavirus proteins showing evidence of evolutionary selective pressure (Chen et al., 2005) (Table 2). Confirming its recent acquisition, presence/absence of the E5 protein was used to demonstrate that the integration of the E5 ORF in the *Deltapapillomavirus* genus occurred between 65 and 23 million years ago (Garcia-Vallve et al., 2005), following the divergence between Camelids and the other ungulates.

It is tempting to speculate that the expansion of the viral arsenal through acquisition of an E6/E7 allowed the proto-papillomavirus to expand its reach into different ecological niches. The E6 and E7 proteins evolved and allowed each virus to become increasingly specialized. However, evolutionary events selected for the loss of these proteins from certain genomes. Likewise, a subset of papillomaviruses acquired an additional "adaptive protein" in the form of E5, providing these viruses with another evolutionary tool.

Role of recombination during papillomavirus phylogenetic evolution

The analysis of the papillomaviral evolutionary history is complicated by the observation of both inter-genetic (i.e., early vs. late) and intra-genetic (e.g. within L2) phylogenetic incongruence (Angulo and Carvajal-Rodriguez, 2007; Bravo and Alonso, 2007; Carvajal-Rodriguez, 2008; Garcia-Vallve et al., 2005; Narechania et al., 2005a). Viral recombination events may provide an explanation for these incongruent phylogenetic trees.

The best-studied examples of recombination events involve papillomaviruses infecting the order Cetacea (whales, dolphins and porpoises; Gottschling et al., 2011a; Rector et al., 2008). When using the late proteins as the base of the phylogenetic tree, these viruses cluster at the root of the *Alphapapillomavirus* genus. However, the early genes suggest a closer relationship to the *Xi-* and *Phipapillomavirus* genera containing viruses that infect cows and goats respectively (Gottschling et al., 2011a). A possible recombination event was mapped near the end of E2/beginning of L2. A second possible recombination site was estimated to lie within the viral LCR (Rector et al., 2008). A virus causing carcinomas in marsupials provides additional evidence for the occurrence of recombination. This virus is likely to be the result of a recombination event between members of the *Papillomaviridae* and *Polyomaviridae* (Woolford et al., 2007).

In addition, computational studies have provided evidence for a handful ancient recombination events, mainly occurring within the L2 ORF (Angulo and Carvajal-Rodriguez, 2007; Carvajal-Rodriguez, 2008; Shah et al., 2010; Varsani et al., 2006).

A recombination event at the root of the *Alphapapillomavirus* genus (Narechania et al., 2005a; Varsani et al., 2006) could potentially have important (clinical) implications. Within this genus, phylogenetic trees inferred from the entire viral genome cluster all cancer-causing types together, suggesting the existence of a common ancestor for the oncogenic human papillomaviruses. However, in separate trees built from the early open reading frames (E1 and E2) or the late ORFs (L1, L2), the carcinogenic potential sorts with the early region of the genome, but not the late region (Fig. 4; Narechania et al., 2005a). A recombination event at the root of the *Alphapapillomavirus* may explain this observation (Varsani et al., 2006). However, since no experimental evidence of recombination between *Alphapapillomaviruses* exists one cannot exclude a role for convergent evolution following niche adaptation (Castoe et al., 2009; Narechania et al., 2005a).

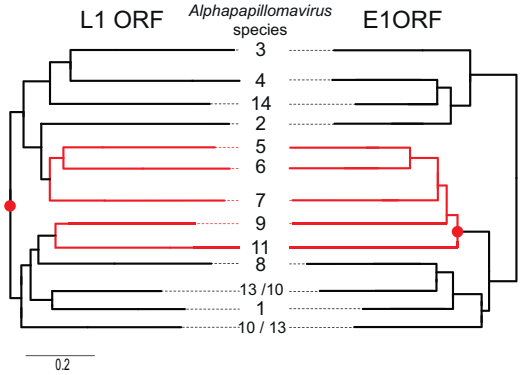

Fig. 4. Phylogenetic incongruence between early and late genes in the *Alphapapillomaviruses*. Phylogenetic trees were inferred based on the L1 (late tree; left) and E1 (early tree; right) using maximum likelihood (Guindon and Gascuel, 2003). Both trees were based on all human viruses within the *Alphapapillomavirus* genus. The different viral species were collapsed for improved readability. The solid branches represent branch-lengths, while the dotted lines were added for ease of interpretation. The viral species are named according to Bernard et al. (2010). Species alpha-12 contains only non-human primate papillomaviruses and was excluded from the analysis. The clade containing the oncogenic viruses is highlighted in red. The red circle illustrates the position of the last common ancestor of the oncogenic viruses within the high-risk clade. This tree shows that oncogenesis maps to the early region, not the late ORFs (Narechania et al., 2005a). The scale bar indicates nucleotide changes per site per year.

The key question is whether a recombination event between two (related) viruses provides the recombinant virus with an (adaptive) advantage over the non-recombinant donor viruses. Through the process of co-evolution, each individual papillomavirus type has optimized its arsenal to successfully infect a specific niche on their host for millions of years. It is therefore unlikely that recombination events will get fixed in the viral population. A recent study provided evidence of several recombination events between two HPV16 variants in a clinical sample (Jiang et al., 2009). However, the recombinant viruses were not detected in the follow-up samples, suggesting that the recombinant viruses had lower fitness compared to the parental viruses. Likewise, a study using experimental infections of rabbits did not provide clear evidence of successful recombination events (Hu et al., 2009).

It is important to consider whether vaccination could provide a selective advantage to specific recombinants. It is conceivable that a recombination event between HPV31 and HPV16 would create a virus carrying HPV16 early (and onco-) genes while coding for the HPV31 structural proteins. Theoretically, this virus would be able to infect the vaccinated population. The pressure applied by vaccination would provide this recombinant with a selective advantage over parental HPV16.

To conclude, while recombination may have played an important role during the evolution of papillomaviruses, it is likely that these events mainly occurred early in the evolutionary process. Two important caveats should be added to this conclusion. Firstly, most methods designed to detect recombination events depend on both donor sequences being present in the database. It is therefore essential that the current efforts in viral discovery be continued. Second, a bias exists in the way samples are collected and chosen for further analysis. In many cases, viruses are detected by PCR primers targeted at a conserved region within the L1 and/or E1 ORF (Rector et al., 2004). The sequence similarity to known viruses in the database determines whether to continue the analysis, thereby potentially missing interesting recombinants.

While the incongruences between the early and late trees are the most striking, it has been shown that all the viral proteins (E1, E2, E6, E7, L1 and L2) evolved (slightly) differently (Garcia-Vallve et al., 2005; Van Doorslaer and Burk, 2010). The observation that phylogenies derived from different proteins are incongruent with each other has several important implications. First, it suggests that evolutionary studies based on "concatenated" data sets (i.e., sequences from different genes are combined into a super-gene) should be interpreted with care. In addition, no single gene tree will accurately represent the evolutionary history of papillomaviruses as a whole, but rather represents the history of the single gene used to construct the tree.

Papillomaviruses evolve about five times faster than their mammalian hosts

In order to estimate the evolutionary rate for fast evolving viruses, it is usually sufficient to obtain the genomic sequence from samples taken at different time points. However, the observation that two isolates of bovine BPV1, collected 30 years apart and from remote cattle populations (Sweden and USA), had nearly identical sequences (Ahola et al., 1983) indicates that papillomaviruses evolve too slowly for this approach. Nevertheless, since virus–host co-speciation is the main evolutionary force involved in shaping the papillomavirus tree (see above), the hosts' divergence times (based on molecular and fossil data) can be used to calibrate a "molecular clock" (Drummond et al., 2006). For example, if the fossil record suggests that two host species diverged at a certain point in time, these hosts' viruses must have diverged at similar times. This approach was used to estimate the evolutionary rate of papillomaviruses infecting carnivores in the *Lambdapapillomavirus* genus (Rector et al., 2007). Based on host calibration points, an evolutionary rate of approximately 1.95×10^{-8} (95% confidence interval 1.32×10^{-8} to 2.47×10^{-8}) nucleotide substitutions per site per year was calculated for the viral coding genome. This rate is comparable with another study which defined an estimate of 4.5×10^{-7} nucleotide substitutions per site per year for the LCR (the most variable region) of the HPV18 genome (Ong et al., 1993). In addition, a recent study based on the complete genome sequence of 108 viruses estimated evolutionary rates of 7.1×10^{-9} (E1 gene) and 9.6×10^{-9} (L1 gene) (Shah et al., 2010) (Table 3).

These slow evolutionary rates suggest that mutations are fixed in the papillomavirus genome at a very low rate. In fact, based on these calculations, papillomaviruses evolve approximately one order of magnitude faster than mammalian host cellular genes, which have a mutation rate of 2.2×10^{-9} nucleotide substitutions per site per year (Kumar and Subramanian, 2002). This is likely due to the high fidelity, proofreading capacity of the host cell DNA replication machinery. In addition, it appears that papillomaviruses are under strong purifying selection pressure (see below), thereby further limiting the fixation of new mutations in the viral genome. In addition, the 5–10-fold increase in mutation rate may be due to the small viral genome size (Drake, 1991; Lynch, 2010) and shorter generation times, resulting in more replication cycles per unit time.

The relative agreement between evolutionary rates derived from different parts of the tree (Table 3) suggests that different parts of the tree evolve at similar rates. Equal evolutionary rates across the papillomavirus tree would allow researchers to apply the precise rates calculated for a specific part of the phylogenetic tree to another part of the tree. For example, the use of the *Lambdapapillomavirus* rates would provide exact estimates of when certain members of the *Alphapapillomavirus* genus acquired the oncogenic phenotype.

A recent study used 108 virus types to estimate the evolutionary rates of the E1 and L1 ORFs (Shah et al., 2010). These rates were slightly slower when compared to the rates reported for the feline papillomaviruses (Rector et al., 2007). However, since mounting evidence suggests that early niche sorting preceded co-evolution, these "whole tree estimates" (as opposed to estimates based on a single genus or clade) may be an underestimate. Nonetheless, the authors concluded that *Alphapapillomaviruses* diverged from their last common ancestor about 75 million years ago. This suggests that the *Alphapapillomavirus* ancestor may have existed prior to the divergence of the primate lineage. A similar timeframe was obtained independently by using HPV18 LCR

Table 3
Papillomavirus evolutionary rates.

	Evolutionary rate	Reference
HPV18 variants		Ong et al. (1993)
URR	4.5×10^{-7}	
Lambdapapillomavirus		Rector et al. (2007)
Overall	1.95E-8 (1.32E-8–2.47E-8)	
E1	1.76E-8 (1.20E-8–2.81E-8)	
E2	2.11E-8 (1.52E-8–2.81E-8)	
E6	2.39E-8 (1.70E-8–3.26E-8)	
E7	1.44E-8 (0.97E-8–2.00E-8)	
L2	2.13E-8 (1.46E-8–2.76E-8)	
L1	1.84E-8 (1.27E-8–2.35E-8)	
URR	2.69E-8 (1.75E-8 –3.69E-8)	
108 Papillomaviruses		Shah et al. (2010)
E1	7.1×10^{-9}	
L1	9.6×10^{-9}	

The mean evolutionary rates (95% confidence interval) in nucleotide substitutions per site per year as calculated for the different papillomavirus genes.

sequences (Ong et al., 1993), validating the conclusion that the divergenece of the human papillomaviruses occurred before the emergence of *Homo sapiens*. Under this hypothesis, it is to be expected that each human *Alphapapillomavirus* species should be mirrored by a non-human primate species. In fact, the observation that papillomaviruses infecting non-human primates are found intertwined with human specific types (as opposed to basal, see above) provides strong support for this idea, and should not be used to dismiss the role of co-evolution.

Additional mechanisms driving papillomavirus evolution

Darwinian evolution

The assumption that papillomaviruses evolve only by genetic drift implies that mutations became fixed at the same rate across the different ORFs. Since this is not the case (see above and Garcia-Vallve et al. (2005) and Rector et al. (2007)), evolutionary forces must differentially affect each individual ORF.

The effect of Darwinian (or diversifying) selection on the evolution of different papillomavirus ORFs has been studied in some detail. The ratio of non-synonymous sites over synonymous sites (dN/dS) estimates the relative importance of selection vs. genetic drift (Kimura, 1977; Yang and Bielawski, 2000). A role for purifying selection is suggested when the dN/dS ratio is smaller than one. Usually these ratios are calculated as an average number across an entire gene. However, given the availability of sufficient sequence date, the selective pressure at a single codon site can be estimated.

To date, limited evidence for positive selection between viral types has been described (Narechania et al., 2005b). An important caveat is that the use of highly divergent sequences saturates the evolutionary changes, thereby interfering with the analysis. However, even at the intra-type level (i.e., by comparing variant genomes), only a few sites have been shown to be under diversifying selection (Table 2). On the contrary, most of the papillomavirus genes are under strong purifying selection, thereby limiting changes to the encoded proteins (Carvajal-Rodriguez, 2008; Chen et al., 2005; DeFilippis et al., 2002). For example, analysis of HPV16 found only seven codons to be under diversifying selection (Carvajal-Rodriguez, 2008; Chen et al., 2005; DeFilippis et al., 2002). It is difficult to provide experimental evidence for the processes involved in directional pressure, however available data allows us to hypothesize about potential sources of pressure. For example, the arginine at position 17 in the HPV16 E6 protein (Arg17) has been shown to be positively selected (Table 2). This residue is involved in the interaction with E6AP (Zanier et al., 2013). However, based on mutagenesis studies, it is unlikely that Arg17 is being selected to stabilize this contact. On the other hand, the cellular immune response may be applying pressure on this Arg17 residue. The naturally occurring HPV16 E6 variant R10G was demonstrated to alter an HLA B*07 binding epitope thereby influencing recognition by cytotoxic T lymphocytes (Ellis et al., 1995). Likewise, the Gln21 and Leu90 have been shown to be located in HLA recognition epitopes (Ellis et al., 1995). This would suggest that the heterogeneity of the host immune response (and the outcome of infection) is actively selecting for specific HPV16 variants.

In summary, it is hard to estimate the impact of Darwinian selection throughout the virus' evolutionary history. However, it appears that interactions with the host are driving the evolution of E5, E6 and E7 proteins (Table 2). This is in agreement with the proposed adaptive role of these proteins (see above). It is plausible that due to their key roles throughout the viral lifecycle, E1, E2, L2 and L1 do not tolerate mutations. The addition of more malleable proteins capable of probing new cellular partners would provide a distinct evolutionary advantage.

Bias in codon usage

Zhou and colleagues made one of the first observations towards the importance of codon usage by papillomavirus genes in 1999 (Zhou et al., 1999). Through changing a number of codons (towards codons more frequently used in humans) the expression of the BPV1 L1 and L2 improved significantly. Furthermore, it was shown that the use of rare tRNAs was indeed rate-limiting for optimal protein translation. This led to the hypothesis that papillomaviruses evolved to use rare codons, thereby limiting the expression of viral proteins, minimizing detection by the immune system.

Indeed, evidence exists that codons complementary to abundant tRNAs provide many lower organisms with an evolutionary advantage (Gouy and Gautier, 1982; Moriyama and Powell, 1997; Powell and Moriyama, 1997; Sharp et al., 1986). However, in mammals there is no clear link between protein abundance and codon usage bias (Sharp et al., 1993). For example, human codon usage bias is determined mainly by genome location and hence mutation pressure acting differently on specific genes. While the basis for the codon bias is not known, its importance during viral replication and gene expression has been suggested (Zhou et al., 1999).

The question remains what processes determine the observed codon bias. The strong correlation between overall GC content and papillomavirus codon bias argues that codon usage is primarily determined by the local dinucleotide content. The viral dinucleotide content is highly similar in genes with different functions and genomic positions. Furthermore, the GC frequencies are similar at synonymous and non-synonymous codon positions. This suggests that genome-wide mutational processes, and not natural selection (which would be expected to act primarily on the non-synonymous codons) are responsible for the observed codon bias in papillomavirus ORFs (Shackelton et al., 2006).

Papillomaviruses have small genomes that are jam packed with open reading frames, many of which are overlapping. In addition to overlapping ORFs, papillomaviruses contain a plethora of non-coding functional elements embedded within ORFs. The presence of these overlapping elements is likely to influence the viral codon usage (Firth and Brown, 2006).

Therefore, no evidence exists to suggest that papillomaviruses "actively" evolved to use sub-optimal codons. Nonetheless, the use of sub-optimal codons may have shaped the evolutionary trajectory of papillomaviruses. It has recently been suggested that codon bias may be aimed at minimizing mutations at the protein level. When hit by a nucleotide mutation, "robust codons" code for amino acids with similar properties. For example, codons beginning with TpA are rare, due to the risk of being mutated to a stop codon (Shackelton et al., 2006). While human genes almost exclusively use "robust codons", almost all papillomavirus codons are biased towards "risky codons" (Bravo and Muller, 2005). This may, in part, provide an explanation for the exceptionally low papillomavirus mutation rate (Duffy et al., 2008) as well as the episodes of rapid radiation (e.g. at the root of the *Betapapillomavirus*) characteristic for the papillomavirus tree. The use of "risky codons" implies that most mutations will prove detrimental for the virus and will therefore not be fixed in the viral genome. The use of "risky codons" may minimize the effects of mutations that may alter the fine balance between the virus and the host reached over millions of years of evolution. However, a rare non-detrimental event would allow the newly minted protein-residue to explore previous unavailable interactions.

While papillomaviruses appear to use sub-optimal codons, the origin and potential (evolutionary) benefit of this adaptation remain elusive.

Evolution of oncogenic risk within the *Alphapapillomavirus* genus

The etiological link between persistent infection with specific oncogenic papillomavirus types and (pre-) cancerous lesions of the cervix has been well established (Schiffman et al., 2007). However, since cellular transformation signals the end of the productive viral infection, papillomaviruses clearly did not evolve to cause cancer. Therefore, a (combination of) viral phenotype(s) essential for the viral lifecycle, create a cellular environment at risk for malignant progression. In order to pinpoint these viral phenotypes, it is essential to gain an understanding into the evolutionary changes that allowed specific papillomavirus access to discreet host ecosystems.

As mentioned above, phylogenetic trees based on different parts of the genome are incongruent. When the early genes are used, all oncogenic *Alphapapillomavirus* types cluster into a monophyletic group as shown in Fig. 4. However, the late genes split oncogenic viruses into two separate clades (Narechania et al., 2005a). This phylogenetic incongruence maps the genetic basis of pathogenicity and oncogenicity to the early genes (Burk et al., 2009).

Phylogenetic analysis based on the early genes group all oncogenic papillomavirus types into a single group. However, not every papillomavirus within this group is actually oncogenic (Bouvard et al., 2010). In addition, a clear spectrum of oncogenic risk exists among the carcinogenic types in the high-risk clade (Schiffman et al., 2009). This would suggest that a careful combination of evolutionary analysis coupled with epidemiological data might be able to pinpoint the underlying genetic changes, especially the observation that certain variants of HPV16 show an increased risk when compared to other variants reinforces this possibility (Schiffman et al., 2010). Nevertheless, the underlying nucleotide changes responsible for the association with cancer have gone largely unsolved (Smith et al., 2011).

In an attempt to understand differences in oncogenic potential, researchers have traditionally compared viral phenotypes between the prototypical high- and low-risk types (HPV16/HPV18 vs. HPV6/HPV11) (Klingelhutz and Roman, 2012). However, these viruses infect different anatomical niches on the human body and are separated by approximately 30 million years of evolutionary changes. It is possible that the observed differences in activity are due to differences in tissue tropism. The interpretation of biochemical data through evolutionary glasses should allow researchers to interpret the confounding effects of evolutionary (niche) adaptation (Van Doorslaer and Burk, 2010). Comparative genomics analyses have been performed to study the impact of distinct viral phenotypes proposed to be important for transformation (Fu et al., 2010; Hiller et al., 2006, 2008; Muench et al., 2009; Van Doorslaer and Burk, 2010; 2012; Van Doorslaer et al., in preparation). The findings in these studies have begun to shed light on which viral phenotypes are correlated with oncogenic potential.

Comparative genomics may provide insights into the viral phenotypes associated with oncogenicity; it does not explain why (papilloma) viruses evolved these phenotypes. It is generally assumed that the viral oncogenes target tumor suppressor pathways in order to usurp the cells' replication machinery to achieve successful viral replication. However, this view may be too simplistic (reviewed in Moore and Chang (2010)). Studies performed in the *Gammaherpesviridae* suggest a role of these pathways in immune evasion as well as viral genome replication. Interestingly, tumor suppressing and innate immunity pathways may share many similarities (Moore and Chang, 2010). For example, both cellular pathways converge on p53 (Takaoka et al., 2003) and p21 (Chin et al., 1996). The anti-antivirus hypothesis (Moore and Chang, 1998, 2003) suggests that viruses target these pathways to interfere with the innate immune system, inadvertently putting the cell at risk for oncogenic progression. While these adaptations would allow papillomaviruses to persist longer in their hosts, the overlap of these cellular functions implies that these cells are also at increased risk for oncogenic transformation.

Impact of vaccines on viral evolution

The highly successful human papillomavirus vaccine targets infection by HPV16 and HPV18 (and HPV6/HPV11, depending on the manufacturer). It is a concern that viruses not targeted by the vaccine would invade the niche vacated through vaccination. Some modeling approaches have suggested that the lack of cross-reactivity (i.e., the vaccine only protects against virus types included, not other related types) will drive the evolution of new oncogenic virus types (Orlando et al., 2012). However, the data summarized in this review clearly shows that papillomaviruses evolve very slowly. In addition, epidemiological studies show that a preexisting infection does not affect the probability of a secondary infection with a related virus (Kaasila et al., 2009; Plummer et al., 2007). Providing further support that papillomaviruses are highly niche specific and do not compete for ecological niches within the epithelium (Stanley et al., 2006).

It is therefore highly unlikely that vaccination will result in the emergence of novel oncogenic viruses in the near future. However, there is little doubt that providing several million years, a virus will eventually emerge to fill this vacated niche. In addition, even in the event of type replacement, the dramatically lower oncogenic potential of the not-included viruses will likely mean that the impact on public health will be manageable.

However, as mentioned above, recombination events may play an important role in vaccine evasion. For example, a recombination event between HPV16 and HPV31 may result in a virus capable of evading the vaccine-mediated immunity, while packing the punch of a highly oncogenic virus. Especially in populations (e.g. HIV+) at risk for infections with multiple viral types (Clifford et al., 2006), the emergence of recombinants should be considered.

Viral evolution in the genomics era

Even though the current papillomavirus sequence databases provide a treasure-trove of sequence data, it is likely that these viruses are only the tip of the iceberg. This is exemplified by the fact that most well studied host species contain a highly diverse set of viruses. However, most amniote hosts have not been sampled, and sampling has been minimal in most other cases. In order to draw general conclusions about papillomavirus evolution we need to fill in the gaps in the phylogenetic tree. The conclusions presented in this review will remain premature until more diverse genomes are discovered and available for analysis.

Just like the advent of PCR and Sanger sequencing provided researchers with the tools to obtain viral sequence data, the rise of pyrosequencing techniques (so called next-gen approaches) is expected to vastly increase the rate of viral discovery. Excitingly, since these approaches do not require prior sequence knowledge, it should be possible to describe drastically novel viral types that would not be detected using the standard consensus primer approach. However, these advances in genome sequencing capabilities should not be used simply to fill our sequence databases with more and more sequence information. If used to address the proper questions, these new methods have the capabilities to illuminate many key issues in (papillomaviral) evolution (Holmes, 2007). We cannot lose track of

the fact that the power of genomics and evolutionary biology is maximized when the genomics data is combined with epidemiological and biochemical data. Therefore, as much as possible, we should attempt to confirm associations with the expected host range and pathology.

Furthermore, an infection may give rise to a "cloud" of (minor) viral populations around the prototype sequence, suggesting mutagenesis of the viral genome inside the host cell. The APOBEC DNA editing system may play a role in this phenomenon (Vartanian et al., 2008). Yet, phylogenetic data provides convincing evidence that papillomaviruses evolve slowly. This suggests that the "cloud" of edited genomes does not get fixed in the population. Next-gen sequencing approaches may allow for a characterization of the viral diversity in a single host (cell). This would provide new insights into the evolutionary dynamics of this exciting family of viruses.

Conclusion

Evolutionary analysis has come a long way since it emerged from the field of systematics. Papillomavirus researchers have a wealth of sequence, epidemiological and biochemical data at their disposal. For these and other reasons, papillomaviruses represent an excellent model system to dissect the role of individual viral nucleotides in all aspects of the viral lifecycle. Building on these arguments, evolutionary analyses will allow researchers to identify those evolutionary changes conferring carcinogenic potential to oncogenic viruses. In addition, evolutionary analysis can, and should, guide the design of mutational studies aimed at understanding the role of a specific protein during the lifecycle.

In short, a thorough understanding of the papillomaviral evolutionary history can and will benefit every aspect of papillomavirus research.

Acknowledgments

I am grateful to Drs. McBride, Khan and Heslin for critical reading of the manuscript.

The author's research is supported by the Intramural Research Program of the NIH, NIAID.

References

Ahola, H., Stenlund, A., Moreno-Lopez, J., Pettersson, U., 1983. Sequences of bovine papillomavirus type 1 DNA: functional and evolutionary implications. Nucleic Acids Res. 11, 2639–2650.
Angulo, M., Carvajal-Rodriguez, A., 2007. Evidence of recombination within human alpha-papillomavirus. Virol. J. 4, 33.
Astell, C.R., Mol, C.D., Anderson, W.F., 1987. Structural and functional homology of parvovirus and papovavirus polypeptides. J. Gen. Virol. 68 (Pt. 3), 885–893.
Bernard, H.U., 1994. Coevolution of papillomaviruses with human populations. Trends Microbiol. 2, 140–143.
Bernard, H.U., 2013. Taxonomy and phylogeny of papillomaviruses: an overview and recent developments. Infect. Genet. Evol..
Bernard, H.U., Burk, R.D., Chen, Z., Van Doorslaer, K., zur Hausen, H., de Villiers, E.M., 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology 401, 70–79.
Bernard, H.U., Chan, S.Y., Delius, H., 1994. Evolution of papillomaviruses. Curr. Top. Microbiol. Immunol. 186, 33–54.
Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., Cogliano, V., 2010. A review of human carcinogens: Part B: biological agents. Lancet Oncol. 10, 321–322.
Bravo, I.G., Alonso, A., 2004. Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth. J. Virol. 78, 13613–13626.
Bravo, I.G., Alonso, A., 2007. Phylogeny and evolution of papillomaviruses based on the E1 and E2 proteins. Virus Genes 34, 249–262.
Bravo, I.G., Muller, M., 2005. Codon usage in papillomavirus genes: practical and functional aspects. Papillomavirus Rep. 16, 63–72.
Brooks, D.R., 2003. The new orthogenesis. Cladistics 19, 443–448.
Brooks, D.R., Ferrao, A.L., 2005. The historical biogeography of co-evolution: emerging infectious diseases are evolutionary accidents waiting to happen. J. Biogeogr. 32, 1291–1299.
Burk, R.D., Chen, Z., Van Doorslaer, K., 2009. Human papillomaviruses: genetic basis of carcinogenicity. Public Health Genomics 12, 281–290.
Carvajal-Rodriguez, A., 2008. Detecting recombination and diversifying selection in human alpha-papillomavirus. Infect. Genet. Evol. 8, 689–692.
Castoe, T.A., de Koning, A.P., Kim, H.M., Gu, W., Noonan, B.P., Naylor, G., Jiang, Z.J., Parkinson, C.L., Pollock, D.D., 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. Proc. Natl. Acad. Sci. U.S.A. 106, 8986–8991.
Chan, S.Y., Bernard, H.U., Ong, C.K., Chan, S.P., Hofmann, B., Delius, H., 1992. Phylogenetic analysis of 48 papillomavirus types and 28 subtypes and variants: a showcase for the molecular evolution of DNA viruses. J.Virol. 66, 5714–5725.
Chan, S.Y., Bernard, H.U., Ratterree, M., Birkebak, T.A., Faras, A.J., Ostrow, R.S., 1997a. Genomic diversity and evolution of papillomaviruses in rhesus monkeys. J.Virol. 71, 4938–4943.
Chan, S.Y., Chew, S.H., Egawa, K., Grussendorf-Conen, E.I., Honda, Y., Rubben, A., Tan, K.C., Bernard, H.U., 1997b. Phylogenetic analysis of the human papillomavirus type 2 (HPV-2), HPV-27, and HPV-57 group, which is associated with common warts. Virology 239, 296–302.
Chan, S.Y., Delius, H., Halpern, A.L., Bernard, H.U., 1995. Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy. J.Virol. 69, 3074–3083.
Chen, Z., Terai, M., Fu, L., Herrero, R., DeSalle, R., Burk, R.D., 2005. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. J. Virol. 79, 7014–7023.
Chen, Z., van Doorslaer, K., Desalle, R., Wood, C.E., Kaplan, J.R., Wagner, J.D., Burk, R.D., 2009. Genomic diversity and interspecies host infection of alpha12 Macaca fascicularis papillomaviruses (MfPVs). Virology 393, 304–310.
Chin, Y.E., Kitagawa, M., Su, W.C., You, Z.H., Iwamoto, Y., Fu, X.Y., 1996. Cell growth arrest and induction of cyclin-dependent kinase inhibitor p21 WAF1/CIP1 mediated by STAT1. Science 272, 719–722.
Clertant, P., Seif, I., 1984. A common function for polyoma virus large-T and papillomavirus E1 proteins? Nature 311, 276–279.
Clifford, G.M., Goncalves, M.A., Franceschi, S., 2006. Human papillomavirus types among women infected with HIV: a meta-analysis. AIDS 20, 2337–2344.
Cole, S.T., Danos, O., 1987. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. J. Mol. Biol. 193, 599–608.
de Villiers, E.M., Fauquet, C., Broker, T.R., Bernard, H.U., zur Hausen, H., 2004. Classification of papillomaviruses. Virology 324, 17–27.
DeFilippis, V.R., Ayala, F.J., Villarreal, L.P., 2002. Evidence of diversifying selection in human papillomavirus type 16 E6 but not E7 oncogenes. J. Mol. Evol. 55, 491–499.
Dobzhansky, T., 1973. Nothing in biology makes sense except in the light of evolution. Am. Biol. Teach. 35, 125–129.
Drake, J.W., 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. U.S.A. 88, 7160–7164.
Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.
Duffy, S., Shackelton, L.A., Holmes, E.C., 2008. Rates of evolutionary change in viruses: patterns and determinants. Nat. Rev. Genet. 9, 267–276.
Ellis, J.R., Keating, P.J., Baird, J., Hounsell, E.F., Renouf, D.V., Rowe, M., Hopkins, D., Duggan-Keen, M.F., Bartholomew, J.S., Young, L.S., et al., 1995. The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. Nat. Med. 1, 464–470.
Firth, A.E., Brown, C.M., 2006. Detecting overlapping coding sequences in virus genomes. BMC Bioinformatics 7, 75.
Fu, L., Van Doorslaer, K., Chen, Z., Ristriani, T., Masson, M., Trave, G., Burk, R.D., 2010. Degradation of p53 by human Alphapapillomavirus E6 proteins shows a stronger correlation with phylogeny than oncogenicity. PLoS One 5, e12816.
Garcia-Vallve, S., Alonso, A., Bravo, I.G., 2005. Papillomaviruses: different genes have different histories. Trends Microbiol. 13, 514–521.
Gottschling, M., Bravo, I.G., Schulz, E., Bracho, M.A., Deaville, R., Jepson, P.D., Van Bressem, M.F., Stockfleth, E., Nindl, I., 2011a. Modular organizations of novel cetacean papillomaviruses. Mol. Phylogenet. Evol. 59, 34–42.
Gottschling, M., Goker, M., Stamatakis, A., Bininda-Emonds, O.R., Nindl, I., Bravo, I.G., 2011b. Quantifying the phylodynamic forces driving papillomavirus evolution. Mol. Biol. Evol. 28, 2101–2113.
Gottschling, M., Kohler, A., Stockfleth, E., Nindl, I., 2007a. Phylogenetic analysis of beta-papillomaviruses as inferred from nucleotide and amino acid sequence data. Mol. Phylogenet. Evol. 42, 213–222.
Gottschling, M., Stamatakis, A., Nindl, I., Stockfleth, E., Alonso, A., Bravo, I.G., 2007b. Multiple evolutionary mechanisms drive papillomavirus diversification. Mol. Biol. Evol. 24, 1242–1258.
Gouy, M., Gautier, C., 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10, 7055–7074.
Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Hafner, M.S., Nadler, S.A., 1988. Phylogenetic trees support the coevolution of parasites and their hosts. Nature 332, 258–259.
Halpern, A.L., 1995. Assessing recombination in HPV, Part I. In: Myers, G., Baker, C. C., Munger, K., Sverdrup, F., McBride, A.A., Bernard, H.U. (Eds.), Human Papillomaviruses. Los Alamos National Laboratory, Los Alamos, NM, p. 1995.

Hiller, T., Poppelreuther, S., Stubenrauch, F., Iftner, T., 2006. Comparative analysis of 19 genital human papillomavirus types with regard to p53 degradation, immortalization, phylogeny, and epidemiologic risk classification. Cancer Epidemiol. Biomarkers Prev. 15, 1262–1267.
Hiller, T., Stubenrauch, F., Iftner, T., 2008. Isolation and functional analysis of five HPVE6 variants with respect to p53 degradation. J. Med. Virol. 80, 478–483.
Holmes, E.C., 2007. Viral evolution in the genomic age. PLoS Biol. 5, e278.
Hu, J., Cladel, N.M., Budgeon, L., Balogh, K.K., Christensen, N.D., 2009. Papillomavirus DNA complementation in vivo. Virus Res. 144, 117–122.
Jiang, M., Xi, L.F., Edelstein, Z.R., Galloway, D.A., Olsem, G.J., Lin, W.C., Kiviat, N.B., 2009. Identification of recombinant human papillomavirus type 16 variants. Virology 394, 8–11.
Kaasila, M., Koskela, P., Kirnbauer, R., Pukkala, E., Surcel, H.M., Lehtinen, M., 2009. Population dynamics of serologically identified coinfections with human papillomavirus types 11, 16, 18 and 31 in fertile-aged Finnish women. Int. J. Cancer 125, 2166–2172.
Kimura, M., 1977. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267, 275–276.
Klingelhutz, A.J., Roman, A., 2012. Cellular transformation by human papillomaviruses: Lessons learned by comparing high- and low-risk viruses. Virology 424, 77–98.
Kumar, S., Subramanian, S., 2002. Mutation rates in mammalian genomes. Proc. Natl. Acad. Sci. U.S.A. 99, 803–808.
Liu, X., Clements, A., Zhao, K., Marmorstein, R., 2006. Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. J. Biol. Chem. 281, 578–586.
Lynch, M., 2010. Evolution of the mutation rate. Trends Genet. 26, 345–352.
Moore, P.S., Chang, Y., 1998. Antiviral activity of tumor-suppressor pathways: clues from molecular piracy by KSHV. Trends Genet. 14, 144–150.
Moore, P.S., Chang, Y., 2003. Kaposi's sarcoma-associated herpesvirus immunoevasion and tumorigenesis: two sides of the same coin? Annu. Rev. Microbiol. 57, 609–639.
Moore, P.S., Chang, Y., 2010. Why do viruses cause cancer? Highlights of the first century of human tumour virology. Nat Rev Cancer 10, 878–889.
Moriyama, E.N., Powell, J.R., 1997. Codon usage bias and tRNA abundance in *Drosophila*. J. Mol. Evol. 45, 514–523.
Muench, P., Hiller, T., Probst, S., Florea, A.M., Stubenrauch, F., Iftner, T., 2009. Binding of PDZ proteins to HPV E6 proteins does neither correlate with epidemiological risk classification nor with the immortalization of foreskin keratinocytes. Virology 387, 380–387.
Narechania, A., Chen, Z., DeSalle, R., Burk, R.D., 2005a. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. J.Virol. 79, 15503–15510.
Narechania, A., Terai, M., Burk, R.D., 2005b. Overlapping reading frames in closely related human papillomaviruses result in modular rates of selection within E2. J. Gen. Virol. 86, 1307–1313.
O'Brien, V., Grindlay, G.J., Campo, M.S., 2001. Cell transformation by the E5/E8 protein of bovine papillomavirus type 4. p27(Kip1), elevated through increased protein synthesis is sequestered by cyclin D1-CDK4 complexes. J. Biol. Chem. 276, 33861–33868.
Ohlenschlager, O., Seiboth, T., Zengerling, H., Briese, L., Marchanka, A., Ramachandran, R., Baum, M., Korbas, M., Meyer-Klaucke, W., Durst, M., Gorlach, M., 2006. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. Oncogene 25, 5953–5959.
Ong, C.K., Chan, S.Y., Campo, M.S., Fujinaga, K., Mavromara-Nazos, P., Labropoulou, V., Pfister, H., Tay, S.K., ter Meulen, J., Villa, L.L., Bernard, H.U., 1993. Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups. J. Virol. 67, 6424–6431.
Orlando, P.A., Gatenby, R.A., Giuliano, A.R., Brown, J.S., 2012. Evolutionary ecology of human papillomavirus: trade-offs, coexistence, and origins of high-risk and low-risk types. J. Infect. Dis. 205, 272–279.
Page, R.D.M., 2002. Tangled Trees: Phylogeny, Cospeciation, and Coevolution, 1st ed. University Of Chicago Press.
Plummer, M., Schiffman, M., Castle, P.E., Maucort-Boulch, D., Wheeler, C.M., 2007. A 2-year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or low-grade squamous intraepithelial lesion. J. Infect. Dis. 195, 1582–1589.
Powell, J.R., Moriyama, E.N., 1997. Evolution of codon usage bias in *Drosophila*. Proc. Natl. Acad. Sci. U.S.A. 94, 7784–7790.
Rebrikov, D.V., Bulina, M.E., Bogdanova, E.A., Vagner, L.L., Lukyanov, S.A., 2002. Complete genome sequence of a novel extrachromosomal virus-like element identified in planarian *Girardia tigrina*. BMC Genomics 3, 15.
Rector, A., Lemey, P., Tachezy, R., Mostmans, S., Ghim, S.J., Van, D.K., Roelke, M., Bush, M., Montali, R.J., Joslin, J., Burk, R.D., Jenson, A.B., Sundberg, J.P., Shapiro, B., Van, R.M., 2007. Ancient papillomavirus-host co-speciation in *Felidae*. Genome Biol. 8, R57.
Rector, A., Stevens, H., Lacave, G., Lemey, P., Mostmans, S., Salbany, A., Vos, M., Van Doorslaer, K., Ghim, S.J., Rehtanz, M., Bossart, G.D., Jenson, A.B., Van Ranst, M., 2008. Genomic characterization of novel dolphin papillomaviruses provides indications for recombination within the *Papillomaviridae*. Virology 378, 151–161.
Rector, A., Van Ranst, M. Animal papillomaviruses. Virology, http://dx.doi.org/10.1016/j.virol.2013.05.007, in press.
Rector, A., Tachezy, R., Van Ranst, M., 2004. A sequence-independent strategy for detection and cloning of circular DNA virus genomes by using multiply primed rolling-circle amplification. J. Virol. 78, 4993–4998.
Rozenblatt-Rosen, O., Deo, R.C., Padi, M., Adelmant, G., Calderwood, M.A., Rolland, T., Grace, M., Dricot, A., Askenazi, M., Tavares, M., Pevzner, S.J., Abderazzaq, F., Byrdsong, M.C., Carvunis, A.R., Chen, A.A., Cheng, J., Correll, M., Duarte, M., Fan, C., Feltkamp, M.C., Ficarro, S.B., Franchi, R., Garg, B.K., Gulbahce, N., Hao, T., Holthaus, A.M., James, R., Korkhin, A., Litovchick, L., Mar, J.C., Pak, T.R., Rabello, S., Rubio, R., Shen, Y., Singh, S., Spangle, J.M., Tasan, M., Wanamaker, S., Webber, J.T., Roecklein-Canfield, J., Johannsen, E., Barabasi, A.L., Beroukhim, R., Kieff, E., Cusick, M.E., Hill, D.E., Munger, K., Marto, J.A., Quackenbush, J., Roth, F.P., DeCaprio, J.A., Vidal, M., 2012. Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature 487, 491–495.
Schiffman, M., Castle, P.E., Jeronimo, J., Rodriguez, A.C., Wacholder, S., 2007. Human papillomavirus and cervical cancer. Lancet 370, 890–907.
Schiffman, M., Clifford, G., Buonaguro, F.M., 2009. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. Infect. Agent Cancer 4, 8.
Schiffman, M., Rodriguez, A.C., Chen, Z., Wacholder, S., Herrero, R., Hildesheim, A., Desalle, R., Befano, B., Yu, K., Safaeian, M., Sherman, M.E., Morales, J., Guillen, D., Alfaro, M., Hutchinson, M., Solomon, D., Castle, P.E., Burk, R.D., 2010. A population-based prospective study of carcinogenic human papillomavirus variant lineages, viral persistence, and cervical neoplasia. Cancer Res. 70, 3159–3169.
Shackelton, L.A., Parrish, C.R., Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evol. 62, 551–563.
Shadan, F.F., Villarreal, L.P., 1995. The evolution of small DNA viruses of eukaryotes: past and present considerations. Virus Genes 11, 239–257.
Shah, S.D., Doorbar, J., Goldstein, R.A., 2010. Analysis of host-parasite incongruence in papillomavirus evolution using importance sampling. Mol. Biol. Evol. 27, 1301–1314.
Sharp, P.M., Stenico, M., Peden, J.F., Lloyd, A.T., 1993. Codon usage: mutational bias, translational selection, or both? Biochem. Soc. Trans. 21, 835–841.
Sharp, P.M., Tuohy, T.M., Mosurski, K.R., 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14, 5125–5143.
Smith, B., Chen, Z., Reimers, L., van Doorslaer, K., Schiffman, M., Desalle, R., Herrero, R., Yu, K., Wacholder, S., Wang, T., Burk, R.D., 2011. Sequence imputation of HPV16 genomes for genetic association studies. PLoS One 6, e21375.
Stanley, M., Lowy, D.R., Frazer, I., 2006. Chapter 12: prophylactic HPV vaccines: underlying mechanisms. Vaccine 24 (3), S3/106–113.
Sundberg, J.P., Van Ranst, M., Burk, R.D., Jenson, A.B., von Krogh, G., Gross, G., 1997. The Nonhuman (Animal) Papillomaviruses: Host Range, Epitope Conservation, and Molecular Diversity, Human Papillomavirus Infections in Dermatovenereology. CRC Press, Boca Raton pp. 47–68.
Takaoka, A., Hayakawa, S., Yanai, H., Stoiber, D., Negishi, H., Kikuchi, H., Sasaki, S., Imai, K., Shibue, T., Honda, K., Taniguchi, T., 2003. Integration of interferon-alpha/beta signalling to p53 responses in tumour suppression and antiviral defence. Nature 424, 516–523.
Van Doorslaer, K., Burk, R.D., 2010. Evolution of human papillomavirus carcinogenicity. Adv. Virus Res. 77, 41–62.
Van Doorslaer, K., Burk, R.D., 2012. Association between hTERT activation by HPV E6 proteins and oncogenic risk. Virology 433, 216–219.
Van Doorslaer, K., Sidi, A.O., Zanier, K., Rybin, V., Deryckere, F., Rector, A., Burk, R.D., Lienau, E.K., van Ranst, M., Trave, G., 2009. Identification of unusual E6 and E7 proteins within avian papillomaviruses: cellular localization, biophysical characterization, and phylogenetic analysis. J. Virol. 83, 8759–8770.
Van Doorslaer, K., Tan, Q., Xirasagar, S., Bandaru, S., Gopalan, V., Mohamoud, Y., Huyen, Y., McBride, A.A., 2013. The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. Nucleic Acids Res. 41, D571–D578.
Varsani, A., van der Walt, E., Heath, L., Rybicki, E.P., Williamson, A.L., Martin, D.P., 2006. Evidence of ancient papillomavirus recombination. J. Gen. Virol. 87, 2527–2531.
Vartanian, J.P., Guetard, D., Henry, M., Wain-Hobson, S., 2008. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science 320, 230–233.
White, E.A., Howley, P.M., 2013. Proteomic approaches to the study of papillomavirus–host interactions. Virology 435, 57–69.
White, E.A., Kramer, R.E., Tan, M.J., Hayes, S.D., Harper, J.W., Howley, P.M., 2012a. Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity. J. Virol. 86, 13174–13186.
White, E.A., Sowa, M.E., Tan, M.J., Jeudy, S., Hayes, S.D., Santha, S., Munger, K., Harper, J.W., Howley, P.M., 2012b. Systematic identification of interactions between host cell proteins and E7 oncoproteins from diverse human papillomaviruses. Proc. Natl. Acad. Sci. U.S.A. 109, E260–E267.
Woolford, L., Rector, A., Van Ranst, M., Ducki, A., Bennett, M.D., Nicholls, P.K., Warren, K.S., Swan, R.A., Wilcox, G.E., O'Hara, A.J., 2007. A novel virus detected in papillomas and carcinomas of the endangered western barred bandicoot (*Perameles bougainville*) exhibits genomic features of both the *Papillomaviridae* and *Polyomaviridae*. J. Virol. 81, 13280–13290.
Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15, 496–503.
Zanier, K., Charbonnier, S., Sidi, A.O., McEwen, A.G., Ferrario, M.G., Poussin-Courmontagne, P., Cura, V., Brimer, N., Babah, K.O., Ansari, T., Muller, I., Stote, R.H., Cavarelli, J., Vande Pol, S., Trave, G., 2013. Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. Science 339, 694–698.
Zhou, J., Liu, W.J., Peng, S.W., Sun, X.Y., Frazer, I., 1999. Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. J. Virol. 73, 4972–4982.