Infection, Genetics and Evolution 67 (2019) 210–221

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Phylogenetic analysis of *Alphapapillomavirus* based on L1, E6 and E7 regions suggests that carcinogenicity and tissue tropism have appeared multiple times during viral evolution

Alexis Rojas-Cruz, Alejandro Reyes-Bermúdez*

Departamento de Biología, Facultad de Ciencias Básicas, Universidad de la Amazonia, Florencia 180002, Colombia

ARTICLE INFO

Keywords:
Alphapapillomavirus evolution
Mutations
Tissue tropism
Cervical cancer

ABSTRACT

Members of the *Alphapapillomavirus* genus are causative agents for cervix cancer and benign lesions in humans. These viruses are classified according to sequence similarities in their L1 region. Yet, viral carcinogenicity has been associated with variations in the proteins encoded by the E6 and E7 genes. In order to relate evolutionary history with origin of carcinogenicity, we performed phylogenetic reconstructions using both nucleotide and predicted amino acid sequences of the L1, E6 and E7 genes. Whilst phylogenetic analysis of L1 reconstructed genus evolutionary history, phylogenies based on E6 and E7 proteins support the idea that mutations at amino acids S/Tx [V/L] (E6) and LxCxE (E7) might be responsible for carcinogenic potential. These findings indicate that virulence within *Alphapapillomavirus* have appeared multiple times during evolution. Our results reveal that oncogenic potential is not a monophyletic clade-specific adaptation but might be the result of positive selection on random mutations occurring on proteins involved in host infection during viral diversification.

1. Introduction

Alphapapillomavirus is a genus of medically important viruses of the *Papillomaviridae* family (Chen et al., 2018). According to “The Papillomavirus Episteme (PaVE), (2018)” database, the genus contains 64 types identified to date. Members infect epithelial cells and are the most common sexually transmitted infectious agent worldwide (Forman et al., 2012). These viruses can cause human cancers of the anus-genital, head and neck regions (de Villiers et al., 2004). Based on oncogenic potential, they are typically classified as low-risk (LR) or high-risk (HR) types (Chiesa et al., 2016). Infection with high-risk genotypes are responsible for approximately 99.7% of cervical cancer, hence becoming a topic of increasing importance in health programs (Walboomers et al., 1999). Worldwide, *Alpha*-PVs-induced carcinoma is the fourth most common cancer in women with an incidence of 528,000 new cases in recent years, and a mortality rate of about 50% especially in developing countries, where 85% of infected women die (Ferlay et al., 2013).

Human papillomaviruses possess a circular double-strand DNA genome of 8 kb in length that encodes eight open reading frames (ORFs). These, are organized into three general regions: 1) an upstream regulatory region (URR) containing elements that control transcription and replication, 2) an early gene region that encodes the E1, E2, E4, E5, E6 and E7 proteins and 3) a late region encodes the L1 and L2 structural proteins (Danos et al., 1982). To promote viral replication, E6 and E7 proteins, target diverse cellular pathways involved in proliferation and inactivation of cell-cycle checkpoints (Doorbar, 2006; Doorbar, 2005). Among all different viral proteins, E6 and E7 gene products are well-conserved in types that causes persistent infections and high-grade lesions (Giuliano et al., 2002; McBride, 2017). E6 protein (P03126) contains two zinc-binding regions CxxC(x)$_{29}$CxxC able to recognized helical motifs LxxLL in target proteins (Huibregtse et al., 1991; Scheffner et al., 1990; Vande Pol and Klingelhutz, 2013). Likewise, in high-risk viruses, E6's carboxyl-terminal have a highly conserved motif that interact with PSD95/DLG1/ZO1 (PDZ) proteins (Delury et al., 2013). E7 proteins (P03129) contain three domains known as CD1, CD2 and CD3 (McLaughlin-Drubin and Münger, 2009). CD1/CD2 regions have homology to a partial part of the conserved Adenovirus E1A's CR1 and CR2 domains and CD3 is a zinc-binding site CxxC(x)$_{29}$CxxC (Chellappan et al., 1992; Münger et al., 2001a; Songock et al., 2017). Both oncoproteins induce changes in the activity of cell cycle regulators and consequently, are shared by high-risk viral types (Howie et al., 2009; McLaughlin-Drubin and Münger, 2009).

Unlike the E6 and E7 proteins, the L1 gene codes for a major capsid protein with the ability to self-assemble spontaneously into virus-like particles (VLPs) (Doorbar et al., 2016). Due to its fundamental role in

---

* Corresponding author.
E-mail address: an.reyes@udla.edu.co (A. Reyes-Bermúdez).

https://doi.org/10.1016/j.meegid.2018.11.008
Received 22 May 2018; Received in revised form 7 November 2018; Accepted 8 November 2018
Available online 17 November 2018

viral-assembly, this gene has been well conserved during the *Alpha* diversification and thus, it is used to classify them inferring evolutionary history (Chan et al., 1995; de Villiers et al., 2004). In fact, the 2004 classification criteria for papillomavirus were defined based on differences in the L1 region. According to this, different genus must share $< 60\%$ identity, species between 60% and 70%, types between 71% and 89% and variants must differ by 1–2% (Bernard et al., 2010; Burk et al., 2013; de Villiers, 2013; Fauquet, 2005). Historically, papillomaviruses have evolved along with birds, reptiles, and fish for over 350 million years adapting to changes in the epithelium of their ancestral hosts (Van Doorslaer, 2013). Molecular clock data suggest that rates of evolution in these viruses is slow with an estimated of (2×10^{-8}–5×10^{-9} nucleotide substitutions/site/year) and are well-adapted to their hosts (García-Vallvé et al., 2005; Rector et al., 2007). This, has led to the general assumption that "host-linked evolution" as well as natural selection have played pivotal roles in papillomavirus diversification (Chen et al., 2017; Gottschling et al., 2007), with over 240 distinct papillomavirus types classified into 37 genera, papillomavirus may be considered the most successful families of vertebrate viruses (Bernard et al., 2010; de Villiers et al., 2004; Van Doorslaer, 2013).

The evolution of the *Alpha* lineage may have occurred 200,000 years ago, probably as a mucosal infecting type in the last common ancestor of apes and humans (Bernard, 2013; Bernard, 1994; Bravo et al., 2010). It has been suggested that this group of viruses have co-evolved in association with pre-human hominid diversifying into types with different tissue tropism and oncogenic potential (Chen et al., 2017; Varsani et al., 2006). Viral diversity observed in the *Alpha*-PVs group have made them interesting subjects for evolutionary studies, however, research has mostly focused, on oncogenic species (Burk et al., 2011; Chen et al., 2013; Chen et al., 2009). In addition, phylogenetic inconsistencies between early and late genes have been identified for some groups in the *Alpha* lineage (Chen et al., 2015; Narechania et al., 2005). These might be explained by gene/marker specific evolutionary histories, as different genes tell different stories (García-Vallvé et al., 2005).

In the present manuscript we have focused on analyzing sequences for 64 genetically distinct papillomaviruses of the *Alpha* genus from the PaVE database (https://pave.niaid.nih.gov) (Van Doorslaer et al., 2013) using a "single-gene" phylogenetic approach. Although whole genome phylogenies could provide a more comprehensive relationship network between *Alpha*-PVs, the resulting topology will not explain how different parts of the genome are being shaped by distinct evolutionary processes. With this in mind, we aimed to compared phylogenies based on L1, E6 and E7 nucleotide (nt), amino acid (aa) and combined (nt/aa) sequences, to understand the extent by which oncogenic potential (high-risk and low-risk forms) is linked to *Alpha*-PVs' evolutionary history. We paid special attention to selection of random mutations in key oncogenic proteins as a mechanism to determine oncogenic phenotype of viral types during the *Alpha* diversification. Although knowledge regarding the cellular mechanisms underlying human papillomavirus infection is considerable, the understanding of its evolutionary history is not. In order to predict and understand carcinogenicity potential and tissue tropism in *Alpha*-PVs, it is necessary to understand the evolutionary processes responsible for the origin of carcinogenicity and tissue preference.

2. Materials and methods

2.1. Nucleotide and protein sequences

In this study, we used complete sequences of 64 *Alpha*-PVs genomes from PaVE (https://pave.niaid.nih.gov) based on sequence registers from Genbank (Van Doorslaer et al., 2013). All sequences were reviewed and compared with those in PaVE. To conduct phylogenetic analysis, we extracted nucleotide sequences for L1, E6 and E7 regions and predicted their amino acid sequences. Genbank and PaVE accession numbers are available in Table 1.

2.2. Sequence alignment and phylogenetic analysis

Full-length nucleotide sequences were codon-based multiple aligned, and then back-translated into nucleotide alignments using a combination of MUSCLE (Edgar, 2004) and TranslatorX (http://translatorx.co.uk) (Abascal et al., 2010). We performed a test of substitution saturation using DAMBE6 (Xia, 2017) and saturation plots indicated that no substitution saturation was found for each alignment. Therefore, all sites for L1, E6 and E7 were used in phylogenetic analyses. Information about tree topology and branch length of each region was adjusted to a series of variables: 1) we used full-length sequences and did not trim our alignments to tools such as Gblocks, TrimAl or Guidance that cause a negative impact on tree inferences and topologies (Jordan and Goldman, 2012; Privman et al., 2012; Tan et al., 2015), 2) tree-construction was implemented using the same algorithm for nucleotide and translated amino acid sequences to reduce accumulated saturations of silent positions (Agosti, 1996; Gottschling et al., 2007; Narechania et al., 2005) and 3) the best-fit evolutionary model was computerized for both nucleotide and amino acid alignments (Arenas, 2015). The statistical selection of best-fit models of nucleotide substitution was performed using jModelTest v2.1.10 based on PHYML 3.0. (Darriba et al., 2012). ProtTest v3.4.2 (Darriba et al., 2011) was used to identify best-fit models for protein alignments. Analysis of the sequences for each criterion was implemented based on corrected Bayesian Information Criterion (BIC).

Phylogenetic analyses were performed with BI methods using MrBayes v3.2 (http://mrbayes.sourceforge.net) (Ronquist and Huelsenbeck, 2003). It is clear that each codon and its associated amino acid evolved together, but this violates the assumption that the sites within the alignment evolve independently. In order to assess both rationales, the Bayesian analysis was carried out using three datasets for each protein-coding gene (nucleotide/amino acid/combined). According to the criterion, the L1, E6 and E7 genes were fitted to a general-time-reversible model with a proportion of invariant sites and gamma distributed rates ($GTR + \Gamma + I$). For E6 and E7 proteins alignment, the ($JTT + F + \Gamma + I$) fixed rate model was used, for L1 protein alignment the best-fit model ($RtREV + F + \Gamma + I$) was selected. Mixed dataset were treated as two partitions by amino acid and nucleotide sequences for each region with the parameters described above. The gamma distribution was approximated using four rate classes.

We carried out five independent runs, each using four Metropolis-coupled Markov Chain Monte Carlo (MCMCMC) (one cold and three heated) for 10,000,000 generations, sampling every 500 generations. Convergence to the stationary distribution was investigated using the diagnostic software Tracer v1.7 (http://tree.bio.ed.ac.uk) (Rambaut, 2018a). Sufficient mixing of the chains was considered to be reached when the average standard deviation of split frequencies was below 0.01 and PSRF (potential scale reduction factor) approach 1.0. After discarding the first 25% samples as burn-in, the consensus trees and Bayesian posterior probability values at nodes were calculated from the posterior distribution at stationary phase. Trees were displayed using FigTree v1.4.2 (http://tree.bio.ed.ac.uk) (Rambaut, 2018b). Sequences from bovine papillomavirus type 1 (d4BPV1) were used as outgroup for all analysis.

2.3. Statistical tree test

To assess confidence and select the best-supported phylogenetic topology, we conducted nucleotide, amino acid and combined trees comparisons using a significance tests under the Maximum Likelihood criterion (Lemey et al., 2009). Our *a priori* hypothesis states that the

Table 1
Characteristics of the papillomavirus used in this study.

Genus and species [a]	HPV type	IARC cancer risk [b]	Tropism	Host	GenBank Accession Numbers	PaVE Accession Numbers	References
Alphapapillomavirus 1	HPV32	Low	Mucosal	*Homo sapiens*	X74475	HPV32REF	(Delius and Hofmann, 1994)
	HPV42	Undefined	Mucosal	*Homo sapiens*	M73236	HPV42REF	(Philipp et al., 1992)
Alphapapillomavirus 2	HPV3	Undefined	Cutaneous	*Homo sapiens*	X74462	HPV3REF	(Delius and Hofmann, 1994)
	HPV10	Undefined	Cutaneous	*Homo sapiens*	X74465	HPV10REF	(Delius and Hofmann, 1994)
	HPV28	Low	Cutaneous	*Homo sapiens*	U31783	HPV28REF	(unpublished data)
	HPV29	Low	Cutaneous	*Homo sapiens*	U31784	HPV29REF	(unpublished data)
	HPV77	Low	Cutaneous	*Homo sapiens*	Y15175	HPV77REF	(Delius et al., 1998)
	HPV78	Low	Cutaneous	*Homo sapiens*	KC138720	HPV78REF	(Silling et al., 2014)
	HPV94	Undefined	Cutaneous	*Homo sapiens*	AJ620211	HPV94REF	(de Villiers et al., 1997)
	HPV117	Undefined	Cutaneous	*Homo sapiens*	GQ246950	HPV117REF	(Köhler et al., 2010)
	HPV125	Undefined	Cutaneous	*Homo sapiens*	FN547152	HPV125REF	(Kovanda et al., 2011)
	HPV160	Undefined	Cutaneous	*Homo sapiens*	AB745694	HPV160REF	(Mitsuishi et al., 2013)
Alphapapillomavirus 3	HPV61	Low	Mucosal	*Homo sapiens*	U31793	HPV61REF	(unpublished data)
	HPV62	Low	Mucosal	*Homo sapiens*	AY395706	HPV62REF	(Fu et al., 2004)
	HPV72	Low	Mucosal	*Homo sapiens*	X94164	HPV72REF	(Völter et al., 1996)
	HPV81	Low	Mucosal	*Homo sapiens*	AJ620209	HPV81REF	(Matsukura and Sugase, 2001)
	HPV83	Low	Mucosal	*Homo sapiens*	AF151983	HPV83REF	(Brown et al., 1999)
	HPV84	Low	Mucosal	*Homo sapiens*	AF293960	HPV84REF	(Terai and Burk, 2001)
	HPV86	Low	Mucosal	*Homo sapiens*	AF349909	HPV86REF	(Terai and Burk, 2001)
	HPV87	Low	Mucosal	*Homo sapiens*	AJ400628	HPV87REF	(Menzo et al., 2001)
	HPV89	Low	Mucosal	*Homo sapiens*	AF436128	HPV89REF	(Terai and Burk, 2002)
	HPV102	Undefined	Mucosal	*Homo sapiens*	DQ080083	HPV102REF	(Narechania et al., 2005)
	HPV114	Undefined	Mucosal	*Homo sapiens*	GQ244463	HPV114REF	(Ekström et al., 2010)
Alphapapillomavirus 4	HPV2	Low	Cutaneous	*Homo sapiens*	X55964	HPV2REF	(Hirsch-Behnam et al., 1990)
	HPV27	Low	Cutaneous	*Homo sapiens*	X74473	HPV27REF	(Delius and Hofmann, 1994)
	HPV57	Low	Cutaneous	*Homo sapiens*	X55965	HPV57REF	(Hirsch-Behnam et al., 1990)
Alphapapillomavirus 5	HPV26	High	Mucosal	*Homo sapiens*	X74472	HPV26REF	(Delius and Hofmann, 1994)
	HPV51	High	Mucosal	*Homo sapiens*	M62877	HPV51REF	(Lungu et al., 1991)
	HPV69	High	Mucosal	*Homo sapiens*	AB027020	HPV69REF	(Kino et al., 2000)
	HPV82	High	Mucosal	*Homo sapiens*	AB027021	HPV82REF	(Kino et al., 2000)
Alphapapillomavirus 6	HPV30	High	Mucosal	*Homo sapiens*	X74474	HPV30REF	(Delius and Hofmann, 1994)
	HPV53	High	Mucosal	*Homo sapiens*	X74482	HPV53REF	(Delius and Hofmann, 1994)
	HPV56	High	Mucosal	*Homo sapiens*	X74483	HPV56REF	(Delius and Hofmann, 1994)
	HPV66	High	Mucosal	*Homo sapiens*	U31794	HPV66REF	(unpublished data)
Alphapapillomavirus 7	HPV18	High	Mucosal	*Homo sapiens*	X05015	HPV18REF	(Cole and Danos, 1987)
	HPV39	High	Mucosal	*Homo sapiens*	M62849	HPV39REF	(Volpers and Streeck, 1991)
	HPV45	High	Mucosal	*Homo sapiens*	X74479	HPV45REF	(Delius and Hofmann, 1994)
	HPV59	High	Mucosal	*Homo sapiens*	X77858	HPV59REF	(Rho et al., 1994)
	HPV68	High	Mucosal	*Homo sapiens*	DQ080079	HPV68REF	(Narechania et al., 2005)
	HPV70	High	Mucosal	*Homo sapiens*	U21941	HPV70REF	(Forslund and Hansson, 1996)
	HPV85	High	Mucosal	*Homo sapiens*	AF131950	HPV85REF	(Chow and Leong, 1999)
	HPV97	Undefined	Mucosal	*Homo sapiens*	DQ080080	HPV97REF	(unpublished data)
Alphapapillomavirus 8	HPV7	Low	Mucosal and cutaneous	*Homo sapiens*	X74463	HPV7REF	(Delius and Hofmann, 1994)
	HPV40	Low	Mucosal and cutaneous	*Homo sapiens*	X74478	HPV40REF	(Delius and Hofmann, 1994)
	HPV43	Undefined	Mucosal and cutaneous	*Homo sapiens*	AJ620205	HPV43REF	(Lörincz et al., 1989)
	HPV91	Low	Mucosal and cutaneous	*Homo sapiens*	AF419318	HPV91REF	(Terai and Burk, 2002)
Alphapapillomavirus 9	HPV16	High	Mucosal	*Homo sapiens*	K02718	HPV16REF	(Seedorf et al., 1985)
	HPV31	High	Mucosal	*Homo sapiens*	J04353	HPV31REF	(Goldsborough et al., 1989)
	HPV33	High	Mucosal	*Homo sapiens*	M12732	HPV33REF	(Cole and Streeck, 1986)
	HPV35	High	Mucosal	*Homo sapiens*	X74477	HPV35REF	(Delius and Hofmann, 1994)
	HPV52	High	Mucosal	*Homo sapiens*	X74481	HPV52REF	(Delius and Hofmann, 1994)
	HPV58	High	Mucosal	*Homo sapiens*	D90400	HPV58REF	(Kirii et al., 1991)
	HPV67	High	Mucosal	*Homo sapiens*	D21208	HPV67REF	(Kirii and Matsukura, 1998)
Alphapapillomavirus 10	HPV6	Low	Mucosal	*Homo sapiens*	X00203	HPV6REF	(Schwarz et al., 1983)
	HPV11	Low	Mucosal	*Homo sapiens*	M14119	HPV11REF	(Dartmann et al., 1986)
	HPV13	Low	Mucosal	*Homo sapiens*	X62843	HPV13REF	(Van Ranst et al., 1992)
	HPV44	Low	Mucosal	*Homo sapiens*	U31788	HPV44REF	(unpublished data)
	HPV74	Low	Mucosal	*Homo sapiens*	AF436130	HPV74REF	(unpublished data)
Alphapapillomavirus 11	HPV34	High	Mucosal	*Homo sapiens*	X74476	HPV34REF	(Delius and Hofmann, 1994)
	HPV73	High	Mucosal	*Homo sapiens*	X94165	HPV73REF	(Völter et al., 1996)
Alphapapillomavirus 13	HPV54	Low	Mucosal	*Homo sapiens*	U37488	HPV54REF	(unpublished data)
Alphapapillomavirus 14	HPV71	Low	Mucosal	*Homo sapiens*	AB040456	HPV71REF	(Matsukura and Sugase, 2001)
	HPV90	Low	Mucosal	*Homo sapiens*	AY057438	HPV90REF	(Terai and Burk, 2002)
	HPV106	Low	Mucosal	*Homo sapiens*	DQ080082	HPV106REF	(Narechania et al., 2005)
Deltapapillomavirus 4	BPV1			*Bos taurus*	X02346	BPV1REF	(Chen et al., 1982)

[a] The table shows division of the *Papillomaviridae* into genera and species, following the classification criteria according to the de Villiers et al. (2004).

[b] The table shows notation of carcinogenic risk levels assigned by the large case-control study conducted by the International Agency for Research on Cancer (IARC) (Muñoz et al., 2003).

single-gene phylogenetic Bayesian inference based from the information of the nucleotide and amino acid sequences separately, explain better *Alpha*-PVs relationships than the combined analysis. For each tested tree, the site-by-site likelihoods were calculated using TREE-PUZZLE v5.3 (Schmidt et al., 2002) under the WAG + Γ + F evolutionary model. We performed the Shimodaira–Hasegawa (SH) test (Shimodaira and Hasegawa, 1999) and an Approximately Unbiased (AU) test (Shimodaira, 2002) as implemented in CONSEL (Shimodaira and Hasegawa, 2001) with a significance limit $p < .05$. The tests implemented in both programs are based on bootstrapping with the resampling estimated log-likelihoods (RELL) method (Kishino et al., 1990).

2.4. Analysis of primary structure of the E6 and E7 proteins

Although the role of E6 and E7's primary structure in human papillomavirus infection is well characterized, the occurrence of random mutations in regions responsible for oncogenic potential during the *Alpha* diversification are not. In order to analyze the role of random mutation determining viral malignity during the *Alpha* evolution, we compared phylogenies based on life history (L1) and oncogenic markers (E6/E7). To identify functional domains in oncogenic proteins we used as a model, E6 and E7 orthologues from the a9HPV16 virus. Structural motifs were identified using PROSITE (de Castro et al., 2006) and domains identified in the Pfam database (Finn et al., 2016) using CLC Genomics Workbench v10.1.1 (CLCBio, Cambridge, MA). For each species group, sequences were aligned separately using CLUSTALW (Thompson et al., 2002) in TranslatorX (http://translatorx.co.uk) (Abascal et al., 2010). The consensus sequences similarity was calculated by using CLUSTALW (Thompson et al., 2002) and multiple sequence alignments (MSA) were performed with Jalview (http://www.jalview.org) analysis tools (Clamp et al., 2004). Sequences were colored in the Jalview–Zappo default style.

3. Results

3.1. Sequence information

To determine phylogenetic relationships within the *Alpha* genus we generated three different datasets: 1) nucleotide alignments, consisting of gene sequences, 2) amino acid alignments, consisting of predicted protein sequences and 3) combined gene and protein alignments, consisting of concatenated sequences. Numbers of characters and their description are provided in Table 2.

3.2. Selection of the best tree

Both AU and SH tests showed that for all markers the combined phylogeny (nt/aa) produced the topology that best explained the datasets as the *a priori* hypothesis was not significantly accepted. As presented in Table 3, "combined trees" have a stronger statistical support than trees originated from single characters and thus, are the most suitable to infer phylogenetic reconstructions. Likewise, combined topologies displayed well-supported trees with deep and recent node rearrangements, contrasting with those of single character that show the lowest proportion of supported nodes and a general lack of support for species-level relationships (Supplementary Fig. S1).

3.3. Phylogenetic analysis reconstructed species-level relationships

Over 200 papillomaviruses have been identified and completely sequenced (Bzhalava et al., 2015). From these, 64 belong to the *Alpha* genus, which contains viral types with distinct tissue preference and ability to induce human cancers. The known biological and clinical properties of viruses studied in this work, are available in Supplementary Table S1. Phylogenetic inferences based on L1, E6, E7 regions were constructed for all genotypes maintaining carcinogenic risk levels notations assigned by the International Agency for Research on Cancer (IARC) (Muñoz et al., 2003). As expected L1 phylogeny was able to reconstruct all viral types as monophyletic groups, reflecting *Alpha*-PVs evolutionary history (Fig. 1). L1 topology suggest that high-risk oncogenic types have appeared multiple times during viral evolution. Moreover, the fact that high-risk types are found in multiple clades across the phylogeny, together with clustering of oncogenic *Alpha*-PVs as monophyletic clades in E6/E7 topologies (Fig. 2), supports the idea that mutations in these genes have occurred multiple times in the genus and are responsible for determining malignancy potential. Our results suggest that oncogenic potential and tissue tropism are not single adaptations with monophyletic origin but on the contrary, have appeared multiple times during the *Alpha* evolution. For viruses without information in the databases, risk assessment was predicted according to their clustering in the E6 and E7 phylogenies. It is necessary to manage this information with discretion as predictions should be clinically tested.

3.4. A high-risk mucosal infecting virus was likely the last common ancestor for Alpha-PVs

As mention before, the concatenated alignment of 1599 nucleotides and of 533 amino acids was selected as the most significant topology for the L1 marker. In general terms, the phylogeny showed high Bayesian posterior probability values (BPP ≥ 93%) with two groups (1, 2) supported by values of 100. The group 1 was the most divergent lineage, composed exclusively of mucosal high-risk genotypes of the α6 species. The group 2 embraced remaining species within eight clades. The first split of this group (lineage A), showed clade I as its most divergent strain. It consisted of mucosal high-risk genotypes of the α5 species (BPP = 100). The fact that both α6 and α5 species are high-risk mucosal infecting types diverging early in the phylogeny, suggests that an ancestral genotype of this characteristics gave rise to current *Alpha*-PVs diversity. The second split in group 2 (lineage B), shows the separation of clade II and lineage C. For the first time we noted a cluster (BPP = 100) of cutaneous low-risk types (α2 species) branching in *Alpha*-PVs evolution. This result indicates a loss of oncogenic potential during the split of these branches. Nonetheless, early branching in lineage C of clade III (BPP = 100), which contains exclusively mucosal high-risk viruses (α7 species) indicates regaining of oncogenic potential. Interestingly, diverse evolutionary scenarios appear to have favored once more the loss of oncogenic potential in lineage D as shown in branching of clade IV and lineages E and F, which grouped mostly species of low pathological risk (clusters V and VI respectively).

While clade V covered members of the α10 species (BPP = 100), clade VI included α1 and α13 species group. These latter sequences clustered confidently 100 out of 100 times. The fact that both clades consisted of types of mucosal low-risk and branched early in their respective lineages, points out that ancestral strains for E and F lineages were most likely low-risk. Despite this, lineage G acquired once more oncogenic potential as shown by the early branching of clade VII that clustered together oncogenic species group α9 and α11. Although members of these species are high-risk strains, they are sister-group to clade VIII, which contains low-risk genotypes that not only infected the mucosal membranes but also have acquired the ability to infect the epidermis (α8 species). As a whole, topology L1 suggests a polyphyletic origin for high-risk oncogenic viruses from a putative ancestral type with high-virulence adapted to infect mucosal membranes. Moreover, transition from high-risk to low-low risk genotypes appears to be a highly dynamic strategy during viral evolution.

3.5. E6 and E7 phylogenies clustered high-risk types as monophyletic

The most significant topology for E6 was built using nucleotide alignments of 663 lengths, concatenated with amino acid alignments of

Table 2
Datasets analyzed in the present study.

Region	Dataset	No. of characters	Description
L1	Nucleotide	1599	L1 gene sequences of 64 *Alpha*-PVs and one bovine papillomavirus 1
	Amino acid	533	L1 protein sequences of 64 *Alpha*-PVs and one bovine papillomavirus 1
	Combined	2132	L1 gene and protein sequences matrix of 64 *Alpha*-PVs and one bovine papillomavirus 1
E6	Nucleotide	663	E6 gene sequences of 64 *Alpha*-PVs and one bovine papillomavirus 1
	Amino acid	221	E6 protein sequences of 64 *Alpha*-PVs and one bovine papillomavirus 1
	Combined	884	E6 gene and protein sequences matrix of 64 *Alpha*-PVs and one bovine papillomavirus 1
E7	Nucleotide	453	E7 gene sequences of 64 *Alpha*-PVs and one bovine papillomavirus 1
	Amino acid	105	E7 protein sequences of 64 *Alpha*-PVs and one bovine papillomavirus 1
	Combined	558	E7 gene and protein sequences matrix of 64 *Alpha*-PVs and one bovine papillomavirus 1

Table 3
AU and SH test results.

	L1			E6			E7		
Tree	ΔlnL [a]	AU [b]	SH [c]	ΔlnL	AU	SH	ΔlnL	AU	SH
Combined (*best tree*)	−5.44	0.691+	0.741+	−5.23	0.682+	0.760+	−0.27	0.768+	0.828+
Amino acid	5.44	0.478+	0.661+	5.23	0.440+	0.590+	0.27	0.413+	0.668+
Nucleotide	32.49	0.193+	0.243+	11.82	0.276+	0.327+	14.23	0.056+	0.078+

Comparison of alternative trees with CONSEL (Shimodaira and Hasegawa, 2001), inferred from separate maximum-likelihood analyses of nine datasets with the WAG + Γ + F evolutionary model. The "+" signs after the values indicate the null hypothesis could not be rejected.
Greater the *p*-value, the greater the probability that the tree was the *best tree* (Efron et al., 1996).

[a] Δlikelihood, differences between the likelihood of a constrained tree and the maximum-likelihood tree.

[b] Approximately Unbiased Test (*p*-values).

[c] Shimodaira–Hasegawa test (*p*-values).

221 positions. E6 topologies were supported by BPP ≥ 68%, displaying two groups (1, 2) with a confidence value of 100 (Fig. 2A). The first group contained exclusively a mucosal low-risk virus (a13HPV54) of the α13 species. The second group clustered together the remaining species within four clades. Clade I was the most divergent group of lineage A, consisting of the mucosal/cutaneous low-risk strains (α8 species) and of mucosal low-risk types of the α10 species (BPP = 79). Clade II (BPP = 85) was an early branch of lineage B grouping two sublineages. The first one contained species lacking oncological risk (α14, α4, α3 species), and the second clustered all cutaneous low-risk genotypes of the α2 species with high BPP values (100). In this topology, clade II emerged as sister-group of lineage C, which grouped low-risk types α1 species (clade III) and all human papillomavirus highly associated with cervical cancer clustered together in the clade IV (α6, α5, α11, α9, α7 species) (BPP = 72). These results agree with the idea that specific amino acid changes in the E6 protein might be related to the oncogenic phenotype in *Alpha*-PVs.

Likewise, the best phylogenetic reconstruction for the E7 region was supported by concatenated data, based on 453 nucleotides and at 105 residues of amino acids. In general, the tree showed robustness supported by BPP values ≥60%. The topology displayed two groups (1, 2) with a statistical value of 100 (Fig. 2B). Whilst cutaneous low-risk types a2HPV125, a2HPV3 of the α2 species were the only members of group 1, the remaining α2 types species were disperse in basal group 2 within lineages A, B, C, D and E. This result indicates high E7 sequence variation among α2 viruses. Lineage F consisted of all remaining *Alpha*-PVs species grouped into five well-supported clusters. The clades: VI (α3, α14, α4 species), VII (α8, α13 species), VIII (α10 species) and IX (α1 species) were the divergent cluster of lineages F, G, H and I respectively and included mucosal low-risk genotypes. Finally, all high-risk viruses were clustered together within clade X (α9, α7, α11, α5, α6 species). As mentioned before, clustering of oncogenic *Alpha*-PVs as monophyletic clades in E6/E7 topologies suggest that positive selection of mutations in these genes might be responsible for determining malignancy potential.

3.6. Viral carcinogenicity has been linked to conserved domains in E6 and E7 proteins

As mentioned before, a9HPV16's E6 and E7 sequences were selected as models to characterize protein primary structure due to the high-oncogenic potential of this species (Ribeiro et al., 2014). Its primary structure is shown in Fig. 3. a9HPV16's E6 protein consists of 158 amino acids with two conserved zinc-binding $CxxC(x)_{29}CxxC$ regions: E6N situated at residues 37-73AA and E6C at 110-146AA. An interdomain of 36 residues link zinc-binding regions. Likewise, its C-terminal end contains conserved amino acids S/Tx[V/L], implicated in the interaction with various PDZ domain-containing proteins. The presence of this motif only in high-risk viruses suggests a critical role in E6-induced cellular immortalization (Ganti et al., 2015) (Fig. 3A). On the other hand, the E7 protein is a much shorter peptide of 98 amino acids. It consists of three conserved regions named CD1, CD2 and CD3. The CD1 and CD2 domains are located in positions 1 to 37AA consisting each of 15 and 22 residues respectively. The CD2 region presents a conserved motif at positions 22-26AA with a high binding affinity to pRb-LxCxE, followed by two conserved serine residues (S31 and S32) susceptible to casein kinase II (CKII) phosphorylation. Binding to pRb-LxCxE is directly linked with the transforming activities of E7, including tumor initiation and induction of genomic instability (Dick and Rubin, 2013). Finally, CD3 is located at the C-terminus end, consisting of a zinc-binding site $CxxC(x)_{29}CxxC$ similar to the ones present in E6 (Fig. 3B).

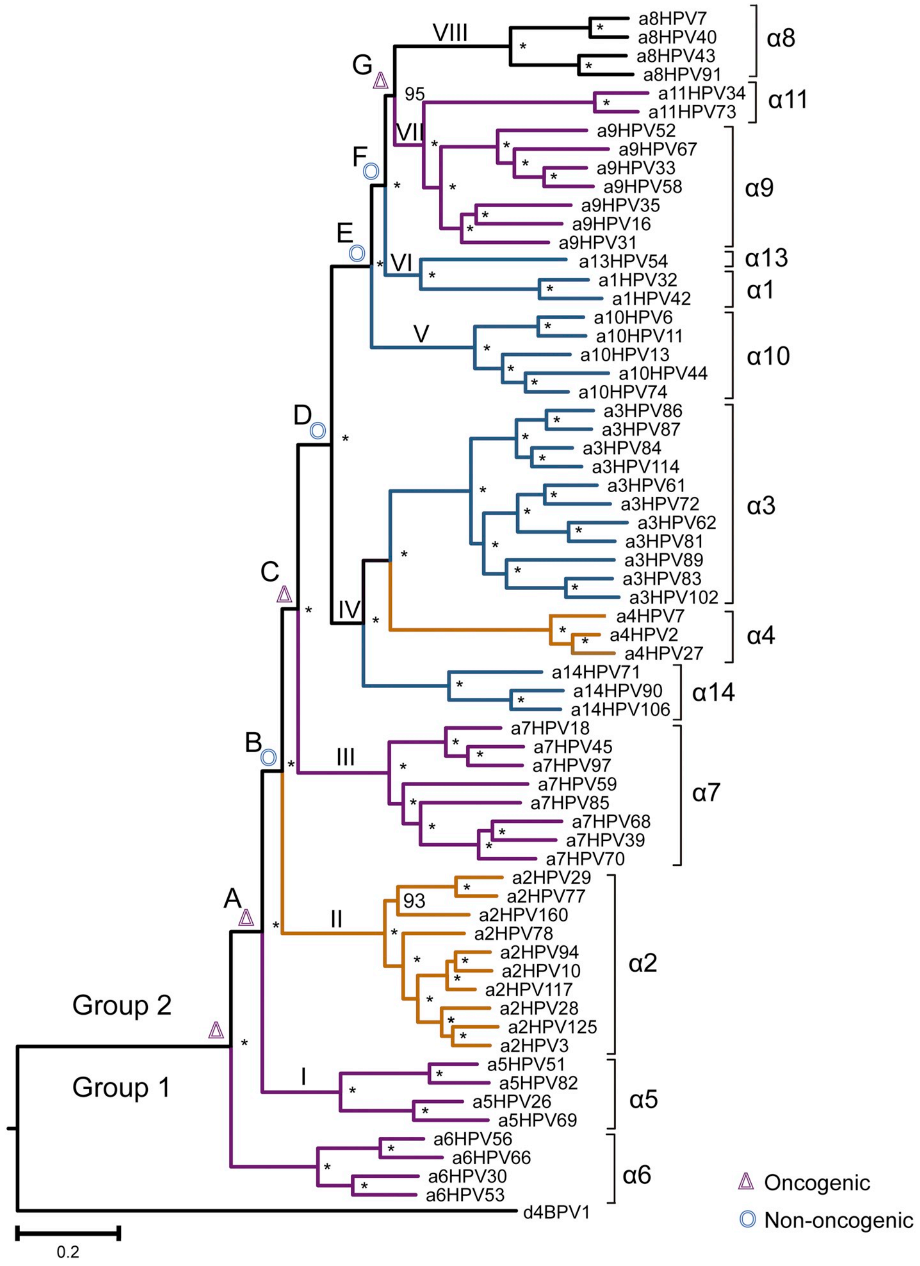

Fig. 1. Bayesian Inference analysis of 64 whole phylogenetically representative *Alpha*-PVs as inferred from a combined data of L1 region justified by AU and SH test (Table 3). Phylogenetic relationships suggest a polyphyletic origin for oncogenic species. Numbers near at nodes indicate Bayesian posterior probability values (BPP). Nodes showing 100% agreement are represented by an asterisk (*). Species groups are identified according to IARC. Mucosal high-risk (purple). Mucosal low-risk (blue). Cutaneous low-risk (brown). Mucosal and cutaneous low-risk (black). Scale bar (bottom left) is proportional to 0.2 substitutions per site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.7. *Mutations in conserved regions of the E6 and E7 proteins are responsible for monophyletic clustering of oncogenic high-risk genotypes*

MSA analysis for E6 and E7 proteins of all *Alpha*-PVs used in this study suggests that variations in these molecules are playing someway in oncogenic phenotype. These results are consistent with description of the protein primary structure of a9HPV16 virus (Fig. 3). MSA of E6 proteins revealed that the C-terminus motif S/Tx[V/L] was well-conserved in high-risk species but not present in low-risk types (Fig. 3A). MSA of E7 proteins showed that the LxCxE binding site and the casein kinase II (SS) within CD2 domain were strongly conserved both for high and low-risk species (Fig. 3B). Interestingly, we identified a high variability of E7-proteins in α2 species, supported by the clustering of α2 as a polyphyletic group. MSA for the E6 and E7 proteins for all 64 genotypes are available in Supplementary Fig. S2.

4. Discussion

4.1. *Carcinogenicity and tissue tropism are not monophyletic traits*

To understand the extent by which oncogenic potential is linked to *Alpha*-PVs' evolution we used publicly available molecular data from all clinically important characterized genotypes. We compared phylogenies based on L1, E6 and E7 nucleotide (nt), amino acid (aa) and combined (nt/aa) sequences and found that topologies based on concatenated sequences had stronger statistical support than those build using single characters (Table 3). Our results are consistent with the idea that including rapid change in silent nucleotide position during phylogenetic reconstruction, results in better-resolved and well-supported topologies (Agosti, 1996). In particular, we paid special attention to mutations in proteins involved in host infection as a mechanism

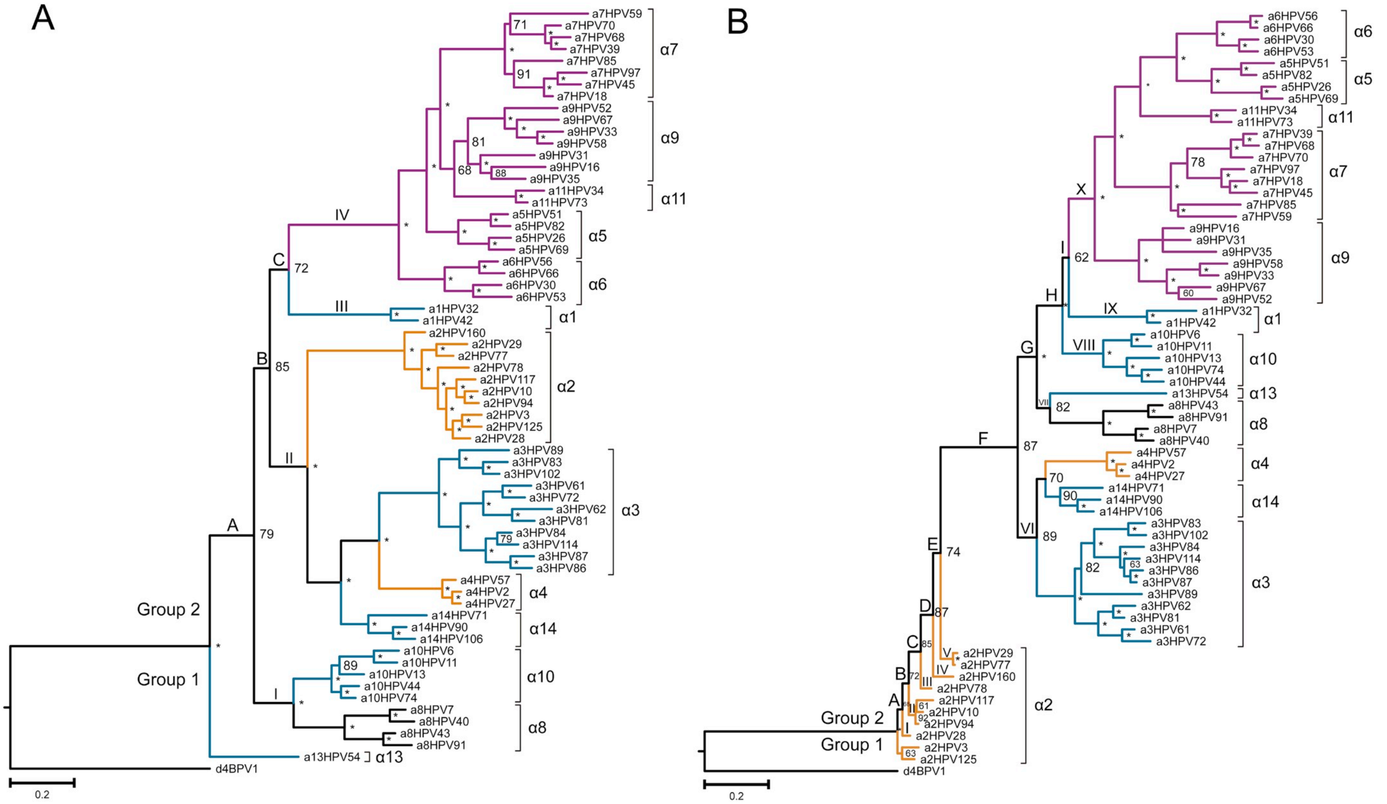

Fig. 2. Bayesian trees based on the analysis of a combined matrix of nucleotides and amino acids of E6/E7 oncoproteins assessed by the AU and SH test, described in Table 3. Mutations in pivotal binding-sites are linked with carcinogenicity status and tissue tropism across phylogenetic *Alpha*-PVs lineages. **(A)**, E6 phylogeny indicates specific amino acids are responsible for the clustering of types viral with high-oncogenic potential associated to mucosal tissue tropism and **(B)**, E7 topology sorts viruses into pathogenicity groups according to affinity LxCxE-pRB binding and reveals high sequence diversity for α2 species. Results from Bayesian inference bootstrap analysis are shown near at nodes and numbers showing 100% agreement are represented by an asterisk (*). Species groups are identified according to IARC. Mucosal high-risk (purple). Mucosal low-risk (blue). Cutaneous low-risk (brown). Mucosal and cutaneous low-risk (black). Scale bar (bottom left) is proportional to 0.2 substitutions per site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to determine oncogenic potential of *Alpha*-PVs types during diversification.

There is evidence showing that whilst early genes E6/E7 have a high divergence rate (García-Vallvé et al., 2005; Hubert, 2005) and are responsible for host infection (Doorbar, 2006). L1 shows a high degree of conservation and thus is a perfect marker to infer evolutionary history (Bernard et al., 2010; Chan et al., 1995; de Villiers et al., 2004). Our findings were consistent with these ideas as: 1) L1 was able to reconstruct all *Alpha*-PVs species independent of their oncogenic potential in a topology likely to reflect ancestral relationships (de Villiers et al., 2004) (Figs. 1) and 2) oncogenic species clustered together in both E6 and E7 phylogenies (Bravo and Alonso, 2004; García-Vallvé et al., 2005; Narechania et al., 2005; Van Doorslaer and Burk, 2010). Although E6 inference was also able to reconstruct all viral species, topologies build from highly conserved sequences are more likely to reflect evolutionary history (Bravo and Alonso, 2007). Interestingly, the E7 marker was also able to reconstruct most *Alpha* species except the α2 group, suggesting that early gene topologies clustered together species according positive selection on a particular trait (Chen et al., 2017) most likely, conserved amino acid positions involved in oncogenic phenotype (Fig. 3).

It is currently accepted that the *Alpha* genus branched early in *Papillomaviridae*, from a mucosal infecting type that targeted the last common ancestor between apes and humans (Bravo et al., 2010). Our results support this hypothesis and suggest that a mucosal high-risk genotype was the last common ancestor of all genus members. The finding that the earliest diverging branch in our L1 topology exclusively contained members of the mucosal high-risk infecting α6 species supports this appreciation. Moreover, the fact that another mucosal high-risk infecting species (α5 species) branched out in the next level of the L1 phylogeny (lineage A) suggests that high-risk carcinogenic potential is an ancestral characteristic and low-risk is a derived trait that appear latter in viral diversification (Fig. 1). It has been proposed that high-oncogenic potential provides an evolutionary advantage by causing disease and thus increased viral fitness (Van Doorslaer, 2013). In a similar way, our L1 results suggest that loss of virulence was to some extent related to colonization of new tissues, originating the cutaneous low-risk infecting α2 species, which appears to be the most ancestral low-risk genotype (Orlando et al., 2012). Virulence loss occurred extensible in the genus during the split of lineage G, evolving the totality of mucosal infecting low-risk genotype known today. During this time, a new skin low-risk infecting genotype also appeared (α4 species). It is not clear whether virulence loss reflects natural selection/genetic drift processes. Oncogenic potential appeared once more in recent viral evolution, giving rise to the mucosa infecting species α9 and α11 (lineage G), which seem to be the most derived *Alpha*-PVs high-risk genotypes. The fact that these two species are sister-group to the mucosal/cutaneous low-risk infecting species α8, suggest that loss of oncogenic potential might be a viral strategy to adapt to new environments during extensive periods of co-speciation (Ong et al., 1993; Shah et al., 2010; Van Doorslaer, 2013). The α8 species is to date, the only known genotype able to infect both mucosa and cutaneous tissues and

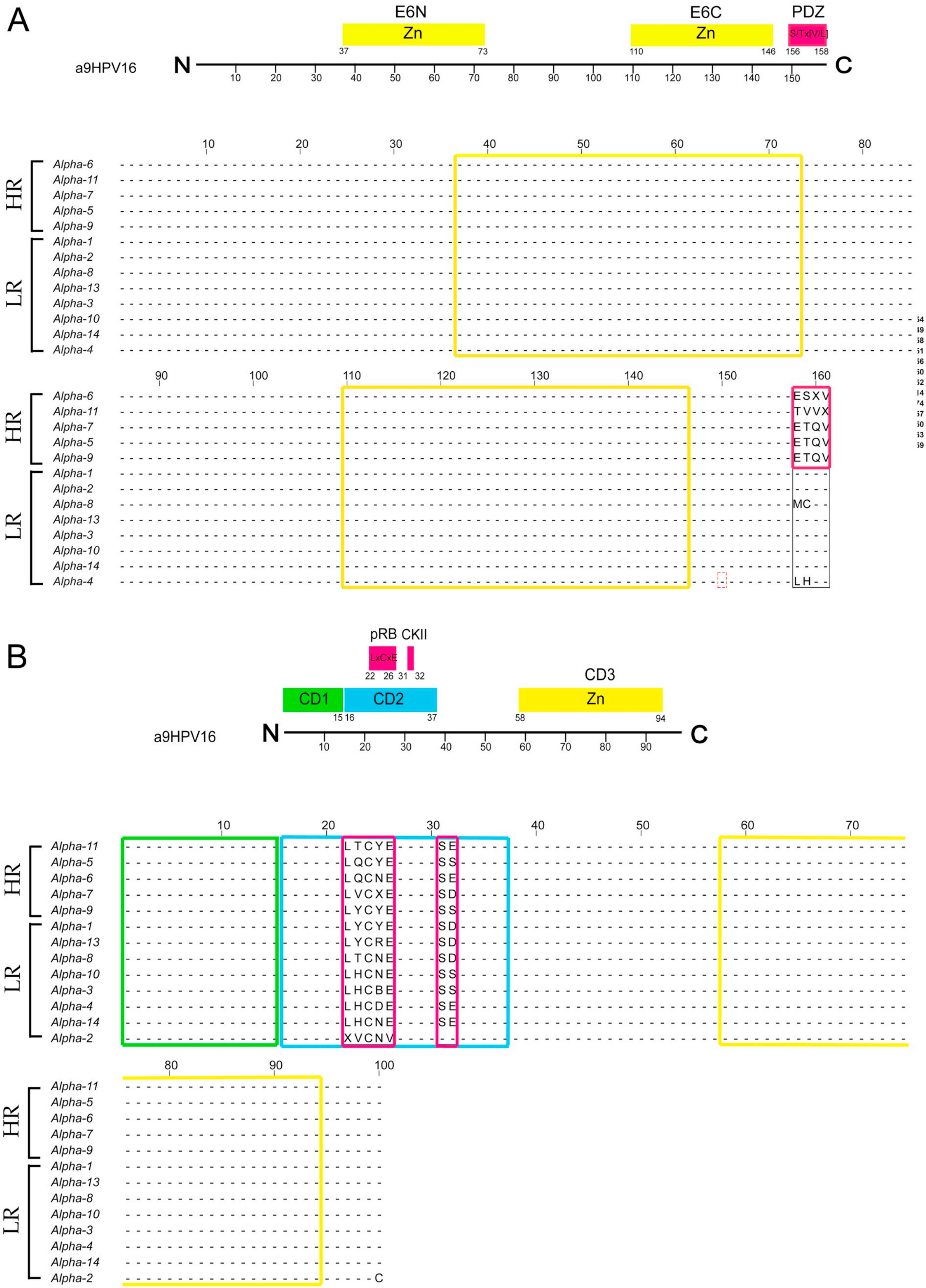

Fig. 3. Primary structure of E6 and E7 oncoproteins from the a9HPV16 virus (top) and MSA of consensus sequence from 14 *Alpha*-PVs species (below). **(A)**, E6 encodes a basic protein of 158 amino acids that contains two zinc-binding regions (yellow) and a S/Tx[V/L] motif (fuchsia) at the C-terminal end, whereas **(B)**, E7 encodes a small protein of 98 amino acids arranged in three conserved regions: CD1 (green), CD2 (blue), and CD3, which contains a zinc-binding region similar to those described in E6 (yellow). CD2 domain (blue) contains a pRb binding-region (LxCxE) (fuchsia) and two serine residues (31S and 32S) susceptible to CKII phosphorylation (fuchsia). Numbers indicate residue positions. HR: High-Risk, LR: Low-Risk. Whilst conserved amino acids are hyphenated, positions with variation are shown by their respective one letter code. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

according to our results belongs to the most recent clade in *Alpha*-PVs evolution (Fig. 1).

4.2. Punctual mutations are playing a role in shaping trees topologies of E6 and E7

As mentioned before, contrasting with L1, E6/E7 oncogenes are highly variable due to point mutations, deletions and insertions (Batista et al., 2011). Despite this, there are conserved regions only present in high-risk viruses that are believed to have roles in viral carcinogenicity (Giuliano et al., 2002; McBride, 2017). There is evidence showing that conserved domains in E6 and E7 proteins, target cellular pathways involved in proliferation and inactivation of cell-cycle checkpoints to promote viral replication (Doorbar, 2006; Doorbar, 2005). Our results are consistent with this idea as topologies for both markers reconstructed oncogenic types as a monophyletic group (Fig. 2). Furthermore, primary structure analysis revealed two conserved zinc-binding regions CxxC(x)$_{29}$CxxC in all E6 proteins as well as a conserve C-terminal region S/Tx[V/L] restricted only to high-risk viruses. While zinc fingers recognized helical motifs LxxLL in target proteins (Huibregtse et al., 1991; Scheffner et al., 1990; Vande Pol and Klingelhutz, 2013), the S/Tx[V/L] motif interacts with PDZ proteins (Delury et al., 2013; Ganti et al., 2015). To some extent, the latter interaction appears to be responsible for oncogenic potential, as: 1) it is responsible to induce cell proliferation in host tissues (Myers and Androphy, 1995), 2) it is restricted only to high-risk viral types, suggesting positive selection for the motif (Pierri et al., 2010; Van Doorslaer et al., 2015) and 3) all S/Tx[V/L] containing viral types cause severe atypical cervical squamous cells and finally carcinoma *in situ* (NIC III) (Hosaka et al., 2013). The fact that all genotypes lacking the S/Tx[V/L] motif, cause small changes in cervical squamous cells but rarely cervical cancer supports this idea (Fig. 3A).

On the other hand, E7 proteins consist of three conserved domains from which, the first 2 (CD1 and CD2) have homology to distinct regions of the E1A Adenovirus and the last one (CD3) is a zinc-binding site CxxC(x)$_{29}$CxxC (Songock et al., 2017). There is evidence demonstrating that E7 conserved domains have roles as cell cycle regulators in host tissues and thus are likely under positive selection in high-risk types (McLaughlin-Drubin and Münger, 2009). The ability of LxCxE motif (within CD2) to interact with pRb proteins disrupting pRb-E2F complexes is perhaps one of the most relevant factors regulating oncogenic level for *Alpha*-PVs (Münger et al., 2001b). This interaction triggers pRb hyper-phosphorylation by cyclin-dependent kinases, inducing premature activation of E2F transcription factor, which activates numerous genes required to enter the S-phase (Liu et al., 2006; Münger et al., 2001a). Consistent with these ideas and similar to E6, the E7 phylogeny clustered together oncogenic genotypes as a monophyletic group, differing only with E6 in the relationships between species. Interestingly, E7 was the only marker unable to reconstruct the α2 species as distinct evolutionary unit (Fig. 2B). Primary structure analysis on E7 proteins showed high sequence variation in the conserved LxCxE sequence within CD2 for α2 species types, result that might explain the phylogeny. As α2 types are cutaneous low-risk infecting viruses is reasonable to think that E7 protein has not direct role determining tissue preference yet, it is likely to regulate oncogenic potential. Contrasting with the E6 MSA analysis where it was easy to identify conserved motifs restricted to oncogenic genotypes, E7 MSA revealed a great degree of conservation in protein structure among *Alpha*-PVs species, excluding α2 species types (Fig. 3B).

At first glance, our results indicate that the primary structure of E7 has not a prominent role determining malignity. However, there is evidence indicating that potential to progress to malignant lesions in oncogenic types is mediated by the efficiency of the LxCxE region to interact with pRb proteins (Dahiya et al., 2000). According to the literature oncogenic viral species cluster in clade X (lineage I) have 10 times more affinity for pRb than remaining viral species (Dick and Dyson, 2002; Heck et al., 1992; McLaughlin-Drubin and Münger, 2009). Although computational approaches have been used to predict protein/domain function (Pierri et al., 2010), experimental work outside the scope of this study is necessary to further test these ideas.

5. Conclusion

All together our results indicate a polyphyletic origin for modern oncogenic high-risk viruses, from a putative mucosal infecting ancestral type with high-virulence. Transition from high-risk to low-risk genotypes appears to be a highly dynamic strategy for viral adaptation. Switches from oncogenic to non-oncogenic types are associated to punctual changes in E6 and E7 oncoproteins that have occurred multiple times during the *Alpha* evolution. Malignancy seems to be mediated by the ability of E6's S/Tx[V/L] motif to modulate cell growth, as well as the ability of the LxCxE motif in E7 proteins to premature initiate S-phase in target cells. Likewise, our results fail to determine the role of the E6 and E7 protein in tissue tropism as: 1) tissue preferences appear to have a polyphyletic origin and 2) protein sequence comparison failed to detect changes restricted to cutaneous/mucosal infecting genotypes. Despite this, loss/regulation of oncogenic potential during evolution appears to have played roles in determining tissue tropism during viral diversification.

Acknowledgments

This study was supported by the Universidad de la Amazonia in Florencia (Caquetá), Colombia.

Author contributions

Alexis Felipe Rojas developed original idea for the study, generated data set, conducted phylogenetic analysis and wrote first draft of manuscript. Alejandro Reyes-Bermudez conceived experimental approach and methodology, analyzed data and contributed materials/analysis tools. All authors wrote the paper.

Conflicts of interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2018.11.008.

References

Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38, W7–W13. https://doi.org/10.1093/nar/gkq291.

Agosti, D., 1996. On combining protein sequences and nucleic acid sequences in phylogenetic analysis: the homeobox protein case. Cladistics 12, 65–82. https://doi.org/10.1006/clad.1996.0005.

Arenas, M., 2015. Trends in substitution models of molecular evolution. Front. Genet. 6 (319). https://doi.org/10.3389/fgene.2015.00319.

Batista, M.V.A., Ferreira, T.A.E., Freitas, A.C., Balbino, V.Q., 2011. An entropy-based approach for the identification of phylogenetically informative genomic regions of papillomavirus. Infect. Genet. Evol. 11, 2026–2033. https://doi.org/10.1016/j.meegid.2011.09.013.

Bernard, H.-U., 1994. Coevolution of papiliomaviruses with human populations. Trends Microbiol. 2, 140–143. https://doi.org/10.1016/0966-842X(94)90602-5.

Bernard, H.-U., 2013. Taxonomy and phylogeny of papillomaviruses: an overview and recent developments. Infect. Genet. Evol. 18, 357–361. https://doi.org/10.1016/j.meegid.2013.03.011.

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. https://doi.org/10.1016/j.virol.2010.02.002.

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. https://doi.org/10.1128/JVI.78.24.13613-13626.2004.

Bravo, I.G., Alonso, Á., 2007. Phylogeny and evolution of papillomaviruses based on the E1 and E2 proteins. Virus Genes 34, 249–262. https://doi.org/10.1007/s11262-006-0017-4.

Bravo, I.G., de Sanjosé, S., Gottschling, M., 2010. The clinical importance of understanding the evolution of papillomaviruses. Trends Microbiol. 18, 432–438. https://doi.org/10.1016/j.tim.2010.07.008.

Brown, D.R., McClowry, T.L., Woods, K., Fife, K.H., 1999. Nucleotide sequence and characterization of human papillomavirus type 83, a novel genital papillomavirus. Virology 260, 165–172. https://doi.org/10.1006/viro.1999.9822.

Burk, R.D., Chen, Z., Harari, A., Smith, B.C., Kocjan, B.J., Maver, P.J., Poljak, M., 2011. Classification and nomenclature system for human Alphapapillomavirus variants: General features, nucleotide landmarks and assignment of HPV6 and HPV11 isolates to variant lineages. Acta Dermatovenerologica Alpina Pannonica, Adriat. 20, 113–123.

Burk, R.D., Harari, A., Chen, Z., 2013. Human papillomavirus genome variants. Virology 445, 232–243. https://doi.org/10.1016/j.virol.2013.07.018.

Bzhalava, D., Eklund, C., Dillner, J., 2015. International standardization and classification of human papillomavirus types. Virology 476, 341–344. https://doi.org/10.1016/j.virol.2014.12.028.

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.

Chellappan, S., Kraus, V.B., Kroger, B., Münger, K., Howley, P.M., Phelps, W.C., Nevins, J.R., 1992. Adenovirus E1A, simian virus 40 tumor antigen, and human papillomavirus E7 protein share the capacity to disrupt the interaction between transcription factor E2F and the retinoblastoma gene product. Proc. Natl. Acad. Sci. U. S. A. 89, 4549–4553.

Chen, E.Y., Howley, P.M., Levinson, A.D., Seeburg, P.H., 1982. The primary structure and genetic organization of the bovine papillomavirus type 1 genome. Nature 299, 529–534.

Chen, Z., Desalle, R., Schiffman, M., Herrero, R., Burk, R.D., 2009. Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97. J. Virol. 83, 1443–1455. https://doi.org/10.1128/JVI.02068-08.

Chen, Z., Schiffman, M., Herrero, R., Desalle, R., Anastos, K., Segondy, M., Sahasrabuddhe, V.V., Gravitt, P.E., Hsing, A.W., Burk, R.D., 2013. Evolution and taxonomic classification of Alphapapillomavirus 7 complete genomes: HPV18, HPV39, HPV45, HPV59, HPV68 and HPV70. PLoS One 8, e72565. https://doi.org/10.1371/journal.pone.0072565.

Chen, Z., de Freitas, L.B., Burk, R.D., 2015. Evolution and classification of oncogenic human papillomavirus types and variants associated with cervical cancer. Methods Mol. Biol. 1249, 3–26. https://doi.org/10.1007/978-1-4939-2013-6_1.

Chen, Z., Ho, W.C.S., Boon, S.S., Law, P.T.Y., Chan, M.C.W., Desalle, R., Burk, R.D., Chan, P.K.S., 2017. Ancient evolution and dispersion of human papillomavirus 58 variants. J. Virol. 91https://doi.org/10.1128/JVI.01285-17. e01285-17.

Chen, Z., Schiffman, M., Herrero, R., Desalle, R., Anastos, K., Segondy, M., Sahasrabuddhe, V.V., Gravitt, P.E., Hsing, A.W., Chan, P.K.S., Burk, R.D., 2018. Classification and evolution of human papillomavirus genome variants: Alpha-5 (HPV26, 51, 69, 82), Alpha-6 (HPV30, 53, 56, 66), Alpha-11 (HPV34, 73), Alpha-13 (HPV54) and Alpha-3 (HPV61). Virology 516, 86–101. https://doi.org/10.1016/j.virol.2018.01.002.

Chiesa, I.J., Perez, M.S., Nuñez, G.G., Pirola, D.A., 2016. Genetic variability and phylogeny analysis of partial L1 gene of human papillomavirus variants in Buenos Aires, Argentina. VirusDisease 27, 41–47. https://doi.org/10.1007/s13337-015-0295-3.

Chow, V.T.K., Leong, P.W.F., 1999. Complete nucleotide sequence, genomic organization and phylogenetic analysis of a novel genital human papillomavirus type, HLT7474-S. J. Gen. Virol. 80, 2923–2929. https://doi.org/10.1099/0022-1317-80-11-2923.

Clamp, M., Cuff, J., Searle, S.M., Barton, G.J., 2004. The Jalview Java alignment editor. Bioinformatics 20, 426–427. https://doi.org/10.1093/bioinformatics/btg430.

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.

Cole, S.T., Streeck, R.E., 1986. Genome organization and nucleotide sequence of human papillomavirus type 33, which is associated with cervical cancer. J. Virol. 58, 991–995.

Dahiya, A., Gavin, M.R., Luo, R.X., Dean, D.C., 2000. Role of the LXCXE binding site in Rb function. Mol. Cell. Biol. 20, 6799–6805.

Danos, O., Katinka, M., Yaniv, M., 1982. Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. EMBO J. 1, 231–236.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: Fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165. https://doi.org/10.1093/bioinformatics/btr088.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772. https://doi.org/10.1038/nmeth.2109.

Dartmann, K., Schwarz, E., Gissmann, L., Zur Hausen, H., 1986. The nucleotide sequence and genome organization of human papilloma virus type 11. Virology 151, 124–130.

de Castro, E., Sigrist, C.J.A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N., 2006. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34, W362–W365. https://doi.org/10.1093/nar/gkl124.

de Villiers, E.-M., 2013. Cross-roads in the classification of papillomaviruses. Virology 445, 2–10. https://doi.org/10.1016/j.virol.2013.04.023.

de Villiers, E.-M., Lavergne, D., McLaren, K., Benton, E.C., 1997. Prevailing papillomavirus types in non-melanoma carcinomas of the skin in renal allograft recipients. Int. J. Cancer 73, 356–361 10.1002/(SICI)1097-0215(19971104)73:3 < 356::AID-IJC9 > 3.0.CO;2-Z.

de Villiers, E.-M., Fauquet, C., Broker, T.R., Bernard, H.-U., Zur Hausen, H., 2004. Classification of papillomaviruses. Virology 324, 17–27. https://doi.org/10.1016/j.virol.2004.03.033.

Delius, H., Hofmann, B., 1994. Primer-directed sequencing of human papillomavirus types. Curr. Top. Microbiol. Immunol. 186, 13–31.

Delius, H., Saegling, B., Bergmann, K., Shamanin, V., de Villiers, E.-M., 1998. The genomes of three of four novel HPV types, defined by differences of their L1 genes, show high conservation of the E7 gene and the URR. Virology 240, 359–365. https://doi.org/10.1006/viro.1997.8943.

Delury, C.P., Marsh, E.K., James, C.D., Boon, S.S., Banks, L., Knight, G.L., Roberts, S., 2013. The role of protein kinase a regulation of the E6 PDZ-binding domain during the differentiation-dependent life cycle of human papillomavirus type 18. J. Virol. 87, 9463–9472. https://doi.org/10.1128/JVI.01234-13.

Dick, F.A., Dyson, N.J., 2002. Three regions of the pRB pocket domain affect its inactivation by human papillomavirus E7 proteins. J. Virol. 76, 6224–6234. https://doi.org/10.1128/JVI.76.12.6224-6234.2002.

Dick, F.A., Rubin, S.M., 2013. Molecular mechanisms underlying RB protein function. Nat. Rev. Mol. Cell Biol. 14, 297–306. https://doi.org/10.1038/nrm3567.

Doorbar, J., 2005. The papillomavirus life cycle. J. Clin. Virol. 32, 7–15. https://doi.org/10.1016/j.jcv.2004.12.006.

Doorbar, J., 2006. Molecular biology of human papillomavirus infection and cervical cancer. Clin. Sci. 110, 525–541. https://doi.org/10.1042/CS20050369.

Doorbar, J., Egawa, N., Griffin, H., Kranjec, C., Murakami, I., 2016. Human papillomavirus molecular biology and disease association. Rev. Med. Virol. 25, 2–23. https://doi.org/10.1002/rmv.1822.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.

Efron, B., Halloran, E., Holmes, S., 1996. Bootstrap confidence levels for phylogenetic trees. Proc. Natl. Acad. Sci. U. S. A. 93, 13429–13434.

Ekström, J., Forslund, O., Dillner, J., 2010. Three novel papillomaviruses (HPV109, HPV112 and HPV114) and their presence in cutaneous and mucosal samples. Virology 397, 331–336. https://doi.org/10.1016/j.virol.2009.11.027.

Fauquet, C., 2005. Virus Taxonomy: Classification and Nomenclature of Viruses: Eighth Report of the International Committee on the Taxonomy of Viruses, 2nd ed. Elsevier Academic Press.

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., et al., 2013. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [WWW Document]. Lyon, Fr. Int. Agency Res. Cancer.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res. 44, D279–D285. https://doi.org/10.1093/nar/gkv1344.

Forman, D., de Martel, C., Lacey, C.J., Soerjomataram, I., Lortet-Tieulent, J., Bruni, L., Vignat, J., Ferlay, J., Bray, F., Plummer, M., Franceschi, S., 2012. Global burden of human papillomavirus and related diseases. Vaccine 30, 12–23. https://doi.org/10.1016/j.vaccine.2012.07.055.

Forslund, O., Hansson, B.G., 1996. Human papillomavirus type 70 genome cloned from overlapping PCR products: complete nucleotide sequence and genomic organization. J. Clin. Microbiol. 34, 802–809.

Fu, L., Terai, M., Matsukura, T., Herrero, R., Burk, R.D., 2004. Codetection of a mixed population of candHPV62 containing wild-type and disrupted E1 open-reading frame in a 45-year-old woman with normal cytology. J. Infect. Dis. 190, 1303–1309. https://doi.org/10.1086/423855.

Ganti, K., Broniarczyk, J., Manoubi, W., Massimi, P., Mittal, S., Pim, D., Szalmas, A., Thatte, J., Thomas, M., Tomaić, V., Banks, L., 2015. The human papillomavirus E6 PDZ binding motif: from life cycle to malignancy. Viruses 7, 3530–3551. https://doi.org/10.3390/v7072785.

García-Vallvé, S., Alonso, A., Bravo, I.G., 2005. Papillomaviruses: different genes have different histories. Trends Microbiol. 13, 514–521. https://doi.org/10.1016/j.tim.2005.09.003.

Giuliano, A.R., Harris, R., Sedjo, R.L., Baldwin, S., Roe, D., Papenfuss, M.R., Abrahamsen, M., Inserra, P., Olvera, S., Hatch, K., 2002. Incidence, prevalence, and clearance of type-specific human papillomavirus infections: the young women's health study. J. Infect. Dis. 186, 462–469. https://doi.org/10.1086/341782.

Goldsborough, M.D., Disilvestre, D., Temple, G.F., Lörincz, A.T., 1989. Nucleotide sequence of human papillomavirus type 31: a cervical neoplasia-associated virus. Virology 171, 306–311.

Gottschling, M., Stamatakis, A., Nindl, I., Stockfleth, E., Alonso, N., Bravo, I.G., 2007. Multiple evolutionary mechanisms drive papillomavirus diversification. Mol. Biol. Evol. 24, 242–258. https://doi.org/10.1093/molbev/msm039.

Heck, D.V., Yee, C.L., Howley, P.M., Munger, K., 1992. Efficiency of binding the retinoblastoma protein correlates with the transforming capacity of the E7 oncoproteins of the human papillomaviruses. Proc. Natl. Acad. Sci. 89, 4442–4446. https://doi.org/10.1073/pnas.89.10.4442.

Hirsch-Behnam, A., Delius, H., de Villiers, E.-M., 1990. A comparative sequence analysis of two human papillomavirus (HPV) types 2a and 57. Virus Res. 18, 81–97. https://doi.org/10.1016/0168-1702(90)90091-O.

Hosaka, M., Fujita, H., Hanley, S.J., Sasaki, T., Shirakawa, Y., Abiko, M., Kudo, M., Kaneuchi, M., Watari, H., Kikuchi, K., Sakuragi, N., 2013. Incidence risk of cervical intraepithelial neoplasia 3 or more severe lesions is a function of human papillomavirus genotypes and severity of cytological and histological abnormalities in adult Japanese women. Int. J. Cancer 132, 327–334. https://doi.org/10.1002/ijc.27680.

Howie, H.L., Katzenellenbogen, R.A., Galloway, D.A., 2009. Papillomavirus E6 proteins.

Virology 384, 324–334. https://doi.org/10.1016/j.virol.2008.11.017.

Hubert, W.G., 2005. Variant upstream regulatory region sequences differentially regulate human papillomavirus type 16 DNA replication throughout the viral life cycle. J. Virol. 79, 5914–5922. https://doi.org/10.1128/JVI.79.10.5914-5922.2005.

Huibregtse, J.M., Scheffner, M., Howley, P.M., 1991. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. EMBO J. 10, 4129–4135.

Jordan, G., Goldman, N., 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. Mol. Biol. Evol. 29, 1125–1139. https://doi.org/10.1093/molbev/msr272.

Kino, N., Sata, T., Sato, Y., Sugase, M., Matsukura, T., 2000. Molecular cloning and nucleotide sequence analysis of a novel human papillomavirus (Type 82) associated with vaginal intraepithelial neoplasia. Clin. Diagn. Lab. Immunol. 7, 91–95.

Kirii, Y., Matsukura, T., 1998. Nucleotide sequence and phylogenetic classification of human papillomavirus type 67. Virus Genes 17, 117–121.

Kirii, Y., Iwamoto, S., Matsukura, T., 1991. Human papillomavirus type 58 DNA sequence. Virology 185, 424–427.

Kishino, H., Miyata, T., Hasegawa, M., 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. 31, 151–160. https://doi.org/10.1007/BF02109483.

Köhler, A., Gottschling, M., Förster, J., Röwert-Huber, J., Stockfleth, E., Nindl, I., 2010. Genomic characterization of a novel human papillomavirus (HPV-117) with a high viral load in a persisting wart. Virology 399, 129–133. https://doi.org/10.1016/j.virol.2009.12.023.

Kovanda, A., Kocjan, B.J., Potočnik, M., Poljak, M., 2011. Characterization of a novel cutaneous human papillomavirus genotype HPV-125. PLoS One 6, e22414. https://doi.org/10.1371/journal.pone.0022414.

Lemey, P., Salemi, M., Vandamme, A.-M., 2009. The phylogenetic handbook. In: Schmidt, H.A. (Ed.), Testing Tree Topologies. Cambridge University Press 2009, United States of America. https://doi.org/10.1017/CBO9780511819049.

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. https://doi.org/10.1074/jbc.M508455200.

Lörincz, A.T., Quinn, A.P., Goldsborough, M.D., Schmidt, B.J., Temple, G.F., 1989. Cloning and partial DNA sequencing of two new human papillomavirus types associated with condylomas and low-grade cervical neoplasia. J. Virol. 63, 2829–2834.

Lungu, O., Crum, C.P., Silverstein, S., 1991. Biologic properties and nucleotide sequence analysis of human papillomavirus type 51. J. Virol. 65, 4216–4225.

Matsukura, T., Sugase, M., 2001. Relationships between 80 human papillomavirus genotypes and different grades of cervical intraepithelial neoplasia: Association and causality. Virology 283, 139–147. https://doi.org/10.1006/viro.2001.0865.

McBride, A.A., 2017. Oncogenic human papillomaviruses. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 372. https://doi.org/10.1098/rstb.2016.0273.

McLaughlin-Drubin, M.E., Münger, K., 2009. The human papillomavirus E7 oncoprotein. Virology 384, 335–344. https://doi.org/10.1016/j.virol.2008.10.006.

Menzo, S., Monachetti, A., Trozzi, C., Ciavattini, A., Carloni, G., Varaldo, P.E., Clementi, M., 2001. Identification of six putative novel human papillomaviruses (HPV) and characterization of candidate HPV type 87. J. Virol. 75, 11913–11919. https://doi.org/10.1128/JVI.75.23.11913-11919.2001.

Mitsuishi, T., Ohsawa, I., Kato, T., Egawa, N., Kiyono, T., 2013. Molecular cloning and characterisation of a novel type of human papillomavirus 160 isolated from a flat wart of an immunocompetent patient. PLoS One 8, e79592. https://doi.org/10.1371/journal.pone.0079592.

Münger, K., Basile, J.R., Duensing, S., Eichten, A., Gonzalez, S.L., Grace, M., Zacny, V.L., 2001a. Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. Oncogene 20, 7888–7898. https://doi.org/10.1038/sj.onc.1204860.

Münger, K., Basile, J.R., Duensing, S., Eichten, A., Gonzalez, S.L., Grace, M., Zacny, V.L., 2001b. Biological activities and molecular targets of the human papillomavirus E7 oncoprotein. Oncogene 20, 7888–7898. https://doi.org/10.1038/sj.onc.1204860.

Muñoz, N., Bosch, F.X., de Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K.V., Snijders, P.J.F., Meijer, C.J.L.M., 2003. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N. Engl. J. Med. 348, 518–527. https://doi.org/10.1056/NEJMoa021641.

Myers, G., Androphy, E., 1995. The E6 protein. [WWW Document]. URL. https://pave.niaid.nih.gov/lanl-archives/compendium/95PDF/3/e6.pdf (accessed 9.25.17).

Narechania, A., Chen, Z., Desalle, R., Burk, R.D., 2005. Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. J. Virol. 79, 15503–15510. https://doi.org/10.1128/JVI.79.24.15503-15510.2005.

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., 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. https://doi.org/10.1093/infdis/jir717.

Philipp, W., Honoré, N., Sapp, M., Cole, S.T., Streeck, R.E., 1992. Human papillomavirus type 42: New sequences, conserved genome organization. Virology 186, 331–334. https://doi.org/10.1016/0042-6822(92)90091-3.

Pierri, C.L., Parisi, G., Porcelli, V., 2010. Computational approaches for protein function prediction: a combined strategy from multiple sequence alignment to molecular docking-based virtual screening. Biochim. Biophys. Acta - Proteins Proteomics 1804, 1695–1712. https://doi.org/10.1016/j.bbapap.2010.04.008.

Privman, E., Penn, O., Pupko, T., 2012. Improving the performance of positive selection inference by filtering unreliable alignment regions. Mol. Biol. Evol. 29, 1–5. https://doi.org/10.1093/molbev/msr177.

Rambaut, A., 2018a. Tracer v1.7. [WWW Document]. URL. http://tree.bio.ed.ac.uk/software/tracer/ (accessed 3.20.18).

Rambaut, A., 2018b. FigTree v1.4.2. [WWW Document]. URL. http://beast.bio.ed.ac.uk/FigTree (accessed 8.28.17).

Rector, A., Lemey, P., Tachezy, R., Mostmans, S., Ghim, S.-J., Van Doorslaer, K., Roelke, M., Bush, M., Montali, R.J., Joslin, J., Burk, R.D., Jenson, A.B., Sundberg, J.P., Shapiro, B., Van Ranst, M., 2007. Ancient papillomavirus-host co-speciation in Felidae. Genome Biol. 8, R57. https://doi.org/10.1186/gb-2007-8-4-r57.

Rho, J., Roy-Burman, A., Kim, H., de Villiers, E.-M., Matsukura, T., Choe, J., 1994. Nucleotide sequence and phylogenetic classification of human papillomavirus type 59. Virology 203, 158–161. https://doi.org/10.1006/viro.1994.1467.

Ribeiro, J., Teixeira, D., Marinho-Dias, J., Monteiro, P., Loureiro, J., Baldaque, I., Medeiros, R., Sousa, H., 2014. Characterization of human papillomavirus genotypes and HPV-16 physical status in cervical neoplasias of women from northern Portugal. Int. J. Gynecol. Obstet. 125, 107–110. https://doi.org/10.1016/j.ijgo.2013.10.011.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Scheffner, M., Werness, B.A., Huibregtse, J.M., Levine, A.J., Howley, P.M., 1990. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell 63, 1129–1136.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18, 502–504. https://doi.org/10.1093/bioinformatics/18.3.502.

Schwarz, E., Dürst, M., Demankowski, C., Lattermann, O., Zech, R., Wolfsperger, E., Suhai, S., Zur Hausen, H., 1983. DNA sequence and genome organization of genital human papillomavirus type 6b. EMBO J. 2, 2341–2348.

Seedorf, K., Krämmer, G., Dürst, M., Suhai, S., Röwekamp, W.G., 1985. Human papillomavirus type 16 DNA sequence. Virology 145, 181–185.

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. https://doi.org/10.1093/molbev/msq015.

Shimodaira, H., 2002. An approximately Unbiased Test of Phylogenetic tree selection. Syst. Biol. 51, 492–508. https://doi.org/10.1080/10635150290069913.

Shimodaira, H., Hasegawa, M., 1999. Multiple Comparisons of Log-Likelihoods with applications to Phylogenetic Inference. Mol. Biol. Evol. 16, 1114–1116. https://doi.org/10.1093/oxfordjournals.molbev.a026201.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246–1247.

Silling, S., Wieland, U., Werner, M., Pfister, H., Potthoff, A., Kreuter, A., 2014. Resolution of novel human papillomavirus-induced warts after HPV vaccination. Emerg. Infect. Dis. 20, 142–145. https://doi.org/10.3201/eid2001.130999.

Songock, W.K., Kim, S., Bodily, J.M., 2017. The human papillomavirus E7 oncoprotein as a regulator of transcription. Virus Res. 231, 56–75. https://doi.org/10.1016/j.virusres.2016.10.017.

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., Dessimoz, C., 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. Syst. Biol. 64, 778–791. https://doi.org/10.1093/sysbio/syv033.

Terai, M., Burk, R.D., 2001. Complete nucleotide sequence and analysis of a novel human papillomavirus (HPV 84) genome cloned by an overlapping PCR method. Virology 279, 109–115. https://doi.org/10.1006/viro.2000.0716.

Terai, M., Burk, R.D., 2002. Identification and characterization of 3 novel genital human papillomaviruses by overlapping polymerase chain reaction: candHPV89, candHPV90, and candHPV91. J. Infect. Dis. 185, 1794–1797. https://doi.org/10.1086/340824.

The Papillomavirus Episteme (PaVE), 2018. Papillomavirus Genome Database. [WWW Document]. URL. https://pave.niaid.nih.gov/#explore/reference_genomes/human_genomes (accessed 6.25.17).

Thompson, J.D., Gibson, T.J., Higgins, D.G., 2002. Multiple sequence alignment using ClustalW and ClustalX. In: Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA. https://doi.org/10.1002/0471250953.bi0203s00. Unit 2.3.

Van Doorslaer, K., 2013. Evolution of the papillomaviridae. Virology 445, 11–20. https://doi.org/10.1016/j.virol.2013.05.012.

Van Doorslaer, K., Burk, R.D., 2010. Evolution of human papillomavirus carcinogenicity. Adv. Virus Res. 77, 41–62. https://doi.org/10.1016/B978-0-12-385034-8.00002-8.

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. https://doi.org/10.1093/nar/gks984.

Van Doorslaer, K., Desalle, R., Einstein, M.H., Burk, R.D., 2015. Degradation of human PDZ-proteins by human Alphapapillomaviruses represents an evolutionary adaptation to a novel cellular niche. PLoS Pathog. 11, e1004980. https://doi.org/10.1371/journal.ppat.1004980.

Van Ranst, M., Fuse, A., Fiten, P., Beuken, E., Pfister, H., Burk, R.D., Opdenakker, G., 1992. Human papillomavirus type 13 and pygmy chimpanzee papillomavirus type 1: Comparison of the genome organizations. Virology 190, 587–596.

Vande Pol, S.B., Klingelhutz, A.J., 2013. Papillomavirus E6 oncoproteins. Virology 445, 115–137. https://doi.org/10.1016/j.virol.2013.04.026.

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. https://doi.org/10.1099/vir.0.81917-0.

Volpers, C., Streeck, R.E., 1991. Genome organization and nucleotide sequence of human papillomavirus type 39. Virology 181, 419–423.

Völter, C., He, Y., Delius, H., Roy-Burman, A., Greenspan, J.S., Greenspan, D., de Villiers, E.-M., 1996. Novel HPV types present in oral papillomatous lesions from patients with HIV infection. Int. J. Cancer 66, 453–456. https://doi.org/10.1002/(SICI)1097-0215(19960516)66:4 < 453::AID-IJC7 > 3.0.CO;2-V.

Walboomers, J.M.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J.F., Peto, J., Meijer, C.J.L.M., Muñoz, N., 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J. Pathol. 189, 12–19. https://doi.org/10.1002/(SICI)1096-9896(199909)189:1 < 12::AID-PATH431 > 3.0.CO;2-F.

Xia, X., 2017. DAMBE6: New tools for microbial genomics, phylogenetics, and molecular evolution. J. Hered. 108, 431–437. https://doi.org/10.1093/jhered/esx033.