Classification and nomenclature system for Human Alphapapillomavirus variants: general features, nucleotide landmarks and assignment of HPV6 and HPV11 isolates to variant lineages

R. D. Burk, Z. Chen, A. Harari, B. C. Smith, B. J. Kocjan, P. J. Maver, M. Poljak

KEY WORDS

human alpha-PV, HPV6, HPV11, HPV variants, classification

ABSTRACT

Background: Papillomaviruses constitute a family of viruses that can be classified into genera, species and types based on their viral genome heterogeneity. Currently circulating infectious human *Alphapapillomaviruses* (alpha-PVs) constitute a set of viral genomes that have evolved from archaic times and display features of host co-speciation. Viral variants are more recently evolved genomes that require a standardized classification and nomenclature.

Objectives: To describe a system for the classification and nomenclature of HPV viral variants and provide landmarks for the numbering of nucleotide positions.

Methods: The complete 8 kb genomes of the alpha-9 species group and HPV6 and 11 types, collected from isolates throughout the world were obtained from published reports and GenBank. Complete genomes for each HPV type were aligned using the E1 start codon and sequence divergence was calculated by global and pairwise alignments using the MUSCLE program. Phylogenetic trees were constructed from the aligned sequences using a maximum likelihood method (RAxML).

Results: Pairwise comparisons of nucleotide differences between complete genomes of each type from alpha-9 HPV isolates (HPV16, 31, 33, 35, 52, 58 and 67) revealed a trimodal distribution. Maximum heterogeneity for variants within a type varied from 0.6%-2.3%. Nucleotide differences of approximately 1.0%-10.0% and 0.5%-1.0% of the complete genomes were used to define variant lineages and sublineages, respectively. Analysis of 43 HPV6 complete genomes indicated the presence of 2 variant lineages, whereas 32 HPV11 isolates were highly similar and clustered into 2 sublineages. A table was constructed of the human alpha-PV landmark nucleotide sequences for future reference and alignments.

Conclusions: A proposed nomenclature system for viral variants and coordination of nucleotide positions will facilitate the comparison of variants across geographic regions and amongst different populations. In addition, this system will facilitate study of pathogenic, tissue tropism and functional differences amongst variant lineages of and polymorphisms within HPV variants.

Introduction

Papillomaviruses are small closed circular double-stranded DNA viruses. They are highly species specific and preferentially infect cutaneous or mucocutaneous epithelium. Papillomavirus genomes have been isolated and characterized from reptiles (1), birds (2), marsupials (3) and multiple other mammalian species (for recent review see (4)) suggesting an evolutionary history spanning more than 300 million years (1). Papillomaviruses replicate their genomes using the host enzymatic machinery, ensuring a high degree of proof reading with low mutation rates (5). Their evolution has been exclusively asexual, although extremely rare recombination events cannot be excluded. This implies that multiple mutations/variations occurring in papillomavirus genomes are not related to genetic distance as in recombining genomes, i.e., linkage disequilibrium, but to sequential accumulation of genetic changes through genetic drift. We have termed this process of speciation through genetic drift and subsequent natural selection, lineage fixation (6). That is, groups of single nucleotide polymorphisms and/or insertions/deletions (indels) tend to become fixed within viral lineages. As time goes on, the quantity of these lineage-defining variations grows eventually leading to speciation.

A distinct human papillomavirus (HPV) "type" is established when the DNA sequence of the L1 open reading frame (ORF) of the cloned viral genome differs from that of any other characterized type by at least 10% (4, 7). Within the PV research community, isolates of the same HPV type are referred to as variants or subtypes when the nucleotide sequences differ by less than 10%. The criterion for HPV types have proven extremely stable and useful for basic researchers, clinicians, epidemiologists and vaccinologists. Nevertheless, the development of a common nomenclature for HPV variants for the multiplicity of HPV types has lagged behind, but is currently being implemented (8).

Over 150 HPV types have been fully characterized; approximately 60 of these are predominantly detected in mucosal epithelia and sort to the Alphapapillomavirus (alpha-PV) genus (4, 7). Human alpha-PV infections are involved in the development of both benign and malignant disease, e.g., condylomata acuminata, respiratory papillomatosis and cervical cancer, respectively. Genital warts, one of the most frequent sexually transmitted infections (STIs) (9), and respiratory/laryngeal papillomas are predominantly caused by HPV6 and HPV11 of the alpha-10 species group (7). Cervical cancer is the most common gynecologic malignancy and one of the leading causes of cancer mortality in women worldwide (10). Most oncogenic or high-risk (HR) types associated with invasive cervical cancer (11-13) are clustered in the HPV alpha-9 species group (14) and account for ~75% of all cervical cancers worldwide (12, 13). Despite phylogenetic relatedness, HPV variants can differ in pathogenicity. For instance, there is a greater risk of cervical cancer for non-European HPV16 variants compared to European variants (15-17).

The establishment of a coherent classification and nomenclature system for HPV variant lineages will facilitate comparison amongst studies that directly determine HPV sequences, as is becoming more common with application of next-generation sequencing methods (18). Moreover, specific variants occur on lineages that are stable and have correlated changes and diagnostic polymorphisms throughout the genome (6, 19). Without a system for naming variant lineages, investigators have had to rely on referring to specific changes at nucleotide positions. This is further complicated by the difficulty in the lack of a system for naming nucleotide positions in each genome.

In this report, we review the evidence for the characterization of variant lineages and sublineages using the largest dataset of complete viral genomes from the medically relevant alpha-9 species group that includes: HPV16, 31, 33, 35, 52, 58 and 67. We extend this nomenclature system to HPV6 and HPV11 that are the main cause of genital warts and laryngeal papillomas. In addition, we present a table of nucleotide landmarks for the human alpha-PV types to facilitate the future description of variants and genotype-phenotype associations.

Materials and Methods

HPV Genome sequences

The DNA sequences for the alpha-9 HPV types were used as reported in previous studies (8, 20). This set of alpha-9 genomes includes 62 complete genomes (CGs) for HPV16, 23 CGs for HPV31, 21 CGs for HPV33, 24 CGs for HPV35, 23 CGs for HPV52, 37 CGs for HPV58 and 8 CGs for HPV67. There were 43 CG sequences for HPV6 and 32 CGs for HPV11. These genomes are accessible through GenBank with the listed names in Figures 2 and 3.

Evolutionary analyses and phylogenetic tree construction

The nucleotide sequences of the complete circular genomes were linearized at the first ATG of the E1 open reading frame (ORF) (see Table 1) and globally aligned using the program MUSCLE (21). The p-distance method in MEGA5 (22) was used to calculate pairwise differences comparing each isolate to all oth-

Table 1. Nucleotide landmarks of human Alphapapillomavirus type genomes.

Species Group	Type	GenBank #	Genome Size	Position of 1st E6 ATG	E6 1st 8bp	1st 8bp of Genome Sequence	Position of 1st E1 ATG	E1 1st 8bp
alpha-1	HPV32*	X74475	7961	102	ATGGCAAG	TAATCTTT	850	ATGGCGGA
alpha-1	HPV42*	M73236	7917	114	ATGTCAGG	CTTATTAT	829	ATGGCGGA
alpha-2	HPV3*	X74462	7820	102	ATGGCAGT	TCTAACTA	806	ATGGATGA
alpha-2	HPV10*	X74465	7919	102	ATGTCCAT	TTATAAAC	791	ATGGACGA
alpha-2	HPV28*	U31783	7959	102	ATGGATGA	TAAATAAT	788	ATGGATGA
alpha-2	HPV29*	U31784	7916	102	ATGTCCAG	TATAAACT	803	ATGGCCGA
alpha-2	HPV77*	Y15175	7887	102	ATGTCTAC	TATAAACT	803	ATGGCTGA
alpha-2	HPV94*	AJ620211	7881	93	ATGTCTAT	TAATGTAG	785	ATGGACGA
alpha-2	HPV117*	GQ246950	7895	103	ATGTCTAT	TTATAAAC	795	ATGGACGA
alpha-2	HPV125^#	FN547152	7809	1	ATGTCTAT	ATGTCTAT	693	ATGGCTGA
alpha-3	HPV61*	U31793	7989	102	ATGGGACC	TAACAATC	811	ATGGCTGA
alpha-3	HPV62^*	AY395706	8092	1	ATGACTGC	ATGACTGC	719	ATGGCCGA
alpha-3	HPV72*	X94164	7988	102	ATGCCTAT	ATTACTAA	832	ATGGCCAA
alpha-3	HPV81*	AJ620209	8070	102	ATGGTCAG	CTTCCTTT	844	ATGGCTGA
alpha-3	HPV83^*	AF151983	8104	1	ATGTCAGG	ATGTCAGG	718	ATGGCGGA
alpha-3	HPV84^*	AF293960	7948	1	ATGCCCAA	ATGCCCAA	715	ATGGCAGA
alpha-3	HPV86^*	AF349909	7983	1	ATGCCCAG	ATGCCCAG	709	ATGGCAGA
alpha-3	HPV87*	AJ400628	7998	87	ATGTGCAA	CAACAATC	890	ATGGTACA
alpha-3	HPV89^*	AF436128	8078	1	ATGCCCGG	ATGCCCGG	721	ATGGCAGA
alpha-3	HPV102^*	DQ080083	8078	1	ATGTCAAG	ATGTCAAG	715	ATGGCACA
alpha-3	HPV114*	GQ244463	8069	213	ATGCCCAC	TGGCTGCG	998	ATGGTACA
alpha-4	HPV2*	X55964	7860	89	ATGCACAC	ATAATGTA	812	ATGGAGGA
alpha-4	HPV27*	X74473	7823	99	ATGCGCAC	TATGTGGT	822	ATGGAGGA
alpha-4	HPV57*	X55965	7861	105	ATGTCTGA	TAATATAT	810	ATGGAGGA
alpha-5	HPV26*	X74472	7855	97	ATGTTCGA	TAACAATT	878	ATGGACTG
alpha-5	HPV51*	M62877	7808	97	ATGTTCGA	AACAATTA	874	ATGGACTG
alpha-5	HPV69*	AB027020	7700	102	ATGTTTCA	CTTTTAAC	886	ATGGACTG
alpha-5	HPV82*	AB027021	7871	102	ATGTTTGA	ATACTTTA	876	ATGGACAG
alpha-6	HPV30*	X74474	7852	102	ATGGCTTT	TGAAAGTT	890	ATGGCGTC
alpha-6	HPV53*	X74482	7856	102	ATGGATCG	GAAAGTAA	892	ATGGCGTC
alpha-6	HPV56*	X74483	7845	102	ATGGAGCC	GAAAGTTT	895	ATGGCGTC
alpha-6	HPV66*	U31794	7824	102	ATGGATTC	GAAAGTTT	895	ATGGCATC
alpha-7	HPV18*	X05015	7857	105	ATGGCGCG	ATTAATAC	914	ATGGCTGA
alpha-7	HPV39*	M62849	7833	107	ATGGCGCG	CTTATAAC	928	ATGGCCAA
alpha-7	HPV45*	X74479	7858	102	ATGGCGCG	AATACTTT	914	ATGGCGGA
alpha-7	HPV59*	X77858	7896	55	ATGGCACG	GTTAAGAC	872	ATGGCCGA
alpha-7	HPV68^*	DQ080079	7822	1	ATGGCGCT	ATGGCGCT	823	ATGGCCAA
alpha-7	HPV70*	U21941	7905	107	ATGGCGCG	CTTATAAC	928	ATGGCCAA
alpha-7	HPV85*	AF131950	7812	105	ATGGCTGA	CTTATACT	920	ATGGCCGA
alpha-7	HPV97^*	DQ080080	7843	1	ATGGCGCG	ATGGCGCG	813	ATGGAAGA
alpha-8	HPV7*	X74463	8027	102	ATGTCTGC	TGTTTAAT	868	ATGGCAGA
alpha-8	HPV40*	X74478	7909	102	ATGTCTGC	TTAATAAC	868	ATGGCAGA
alpha-8	HPV43*	AJ620205	7975	102	ATGACTGC	CTAACAAT	835	ATGGCTGA
alpha-8	HPV91^*	AF419318	7966	1	ATGAGTAA	ATGAGTAA	908	ATGGCTGA
alpha-9	HPV16*	K02718	7904	83	ATGCACCA	ACTACAAT	865	ATGGCTGA
alpha-9	HPV16R	⌘	7906	83	ATGCACCA	ACTACAAT	865	ATGGCTGA
alpha-9	HPV31*	J04353	7912	108	ATGTTCAA	TAATAATA	862	ATGGCTGA
alpha-9	HPV33*	M12732	7909	109	ATGTTTCA	GTAAACTA	879	ATGGCCGA
alpha-9	HPV35*	M74117	7851	110	ATGTTTCA	CCCTATAA	868	ATGGCTGA
alpha-9	HPV52*	X74481	7942	102	ATGTTTGA	TAAATTAT	864	ATGGAGGA
alpha-9	HPV58*	D90400	7824	110	ATGTTCCA	CTAAACTA	883	ATGGATGA
alpha-9	HPV67*	D21208	7801	102	ATGTTTCA	TTATAATC	875	ATGGAGGA
alpha-10	HPV6*	X00203	7902	102	ATGGAAAG	GTTAATAA	832	ATGGCGGA
alpha-10	HPV11*	M14119	7931	102	ATGGAAAG	CTTAATAA	832	ATGGCGGA
alpha-10	HPV13*	X62843	7880	104	ATGGAAAG	GTTTCTAA	843	ATGGCAGA
alpha-10	HPV44*	U31788	7833	105	ATGGAAAG	TTAATAAT	832	ATGGCTGA
alpha-10	HPV74^*	AF436130	7887	1	ATGGAAAG	ATGGAAAG	721	ATGGCGGA
alpha-11	HPV34*	X74476	7723	102	ATGTTTTT	ACTATAAT	851	ATGGCTGA
alpha-11	HPV73*	X94165	7700	102	ATGCTGTT	ACTATAAT	850	ATGGCTGA
alpha-13	HPV54*	U37488	7759	12§	ATGATTTA	TAACTACA	828	ATGGCGGA
alpha-14	HPV71*	AB040456	8017	102	ATGCTTGG	TTGTTCTA	838	ATGGCCGA
alpha-14	HPV90^*	AY057438	8033	1	ATGACCAA	ATGACCAA	725	ATGGCCGA
alpha-14	HPV106^*	DQ080082	8035	1	ATGGGTAC	ATGGGTAC	761	ATGGCCGA

^ Corresponds to reference sequence with "A" of the first E6 ATG as position 1
** Nucleotide positions based on genome reference sequence from PaVE (http://pave.niaid.nih.gov)*
 # Nucleotide positions based on genome sequence available from GenBank (http://www.ncbi.nlm.nih.gov)
§ Denotes disagreement between reference genome positions in PaVE and GenBank, PaVE positions are shown
⌘ HPV16R is a revised "virtual" sequence as discussed on page III-117, 1997 HPV Compendium (http://pave.niaid.nih.gov/lanl-archives/compendium/97PDF/3/Meissner.pdf)

er variants of the same type based on the global alignment. The assignment of position numbers for each nucleotide is based on the nucleotide numbering of the prototype reference sequence as shown in Table 1.

Maximum likelihood trees for HPV6 and HPV11 aligned genomes were constructed using RAxML MPI v7.2.8 (23). The GTR + gamma model was set for among-site rate variation and allowed substitution rates of aligned sequences to be different.

Human alpha-PV nucleotide landmarks

The complete genome sequences of 62 reference or prototype human Alphapapillomavirus types were obtained from GenBank or the PaVE website. The circular viral genomes were linearized and aligned based on the 1st ATG site of the E1 ORF using the global alignment software MUSCLE (21). The position of the 1st nucleotide of the E1 ORF start codon, ATG, is given for the prototype reference sequence for each type grouped by species (see Table 1, "Position of 1st E1 ATG"). The genomic position of the 1st nucleotide of the E6 ORF ATG for each of the 62 human alpha-PV types is displayed in Table 1 and the first 8 nucleotides 3' to this site are listed in the neighboring column, "E6 1st 8 bp". Also listed are the first 8 bp of the reference/prototype genomes from the published reports.

Results

HPV variant lineage classification and nomenclature

To establish an unbiased distribution of the relatedness of variant genomes within a given type, we have used the dataset for HPV isolates (HPV16, HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67) from the alpha-9 species group as an example (8, 20). The distribution of percent differences between variants revealed a trimodal pattern (Figure 1A). This trimodal distribution of pairwise comparisons indicates that some variants are more closely related to one another than others, thus supporting a grouping of lineages for each type. Previous examination of phylogenies for each of the alpha-9 types combined with an approximate cut-off of 1.0% difference between genomes was used to define major variant lineages (6, 8). Each major lineage was named using an alphanumeric, with the "A" clade always containing the reference genome for each type. Support for this distinction between variants was examined by viewing the distribution of pairwise comparisons within each variant lineage (i.e., intra-lineage) or between variant lineages for each of the seven HPV types (i.e., inter-lineage), again this analysis only compares isolate genomes within a specific type and summarizes the data for all the individual types (Figure 1B). Although there is a bimodal distribution seen within the inter-lineage comparisons driven by the deeper nodes separating HPV16 (European vs. non-European lineages) (20) and HPV52 (A,B,C vs. D lineages) variants (8), we have not made a distinction at this level of variant divergence. The overlap between the inter- and intra-lineage distributions (0.7%-0.9%) indicates a fixed value cannot be used to distinguish variant lineages. We conservatively suggest a 1.0% divergence, with the caveat that no classification system can exactly categorize the process of evolution. Two distributions were discernable between and within the genome comparisons of sublineages for each HPV type (Figure 1C). Differences between genomes in the 0.5%-1% range were designated as sublineages (e.g., A1, A2, etc.). We have used these criteria to classify HPV6 and HPV11 based on the available genomes (24, 25).

Nomenclature of HPV6 variant isolates

Forty-three HPV6 complete genome sequences were available for analyses. These genomes were characterized from seven isolates from laryngeal papillomas (LP5, LP26, LP130, LP98(131), LP137, LP96(175) and LP11) and 11 isolates from condyloma acuminatum lesions (CAC377, CAC26, CAC251c, CAC306, CAC11, CAC23z, CAC231, CAC96, CAC331, CAC301 and CAC56) obtained from a study of Kocjan et al. (GenBank Accession Numbers: FR751320 – FR751338) (24), and from three previously characterized HPV6 isolates, including prototype HPV6b (X00203), HPV-6vc (AF092932), and HPV6a (L42216). The rest of the HPV6 isolates were obtained from the ongoing research of the HPV6 genomic diversity in Slovenia. As shown in Figure 2, the topology of the tree constructed with HPV6 variants revealed two distinct lineages, termed A and B. The isolates sorting to the A lineage were highly related, although two clades were present differing by ~ 0.2% (Figure 2, right panel). Lineage B was more variable and was further divided into three sublineages B1, B2 and B3, with inter-sublineage differences of 0.4% - 0.7%. These three sublineages were equally distant from the A lineage, with a difference of approximately 1.5% of nucleotide sequences (Figure 2, right panel).

Nomenclature of HPV11 variant isolates

Complete genome sequences were available for 32 HPV11 isolates representing a heterogeneous set of 10 isolates from Slovenian patients with exophytic genital lesions, laryngeal papillomatosis and cervical samples (CS20, A86, A346, CAC86, LP12, CAC246, A48,

LP13, A47 and A260) obtained from a study of Maver et al. (GenBank Acc. Nos: FN870021, FN870022 and FN907957 – FN907964) (25). In addition, genomes of six Hungarian HPV11 isolates from recurrent respiratory papillomatosis (HUNG1, RRP1 to RRP5; GenBank Acc. Nos: FR872717 and HE574701 – HE574705), as well as two previously deposited genomes (prototype HPV11 (M14119) and LZod45-11 from a cervical swab (EU918768) were also available in sequence repositories (26, 27). The rest of the HPV11 isolates were from the ongoing research of the HPV11 genomic diversity in Slovenia. (Figure 3). Nevertheless, all variants were highly conserved; the maximum pairwise difference was approximately 0.4% (Figure 3). Based on the nucleotide difference of the aligned complete genomes and the topology of the phylogenetic tree, we have designated two clades as sublineages A1 and A2.

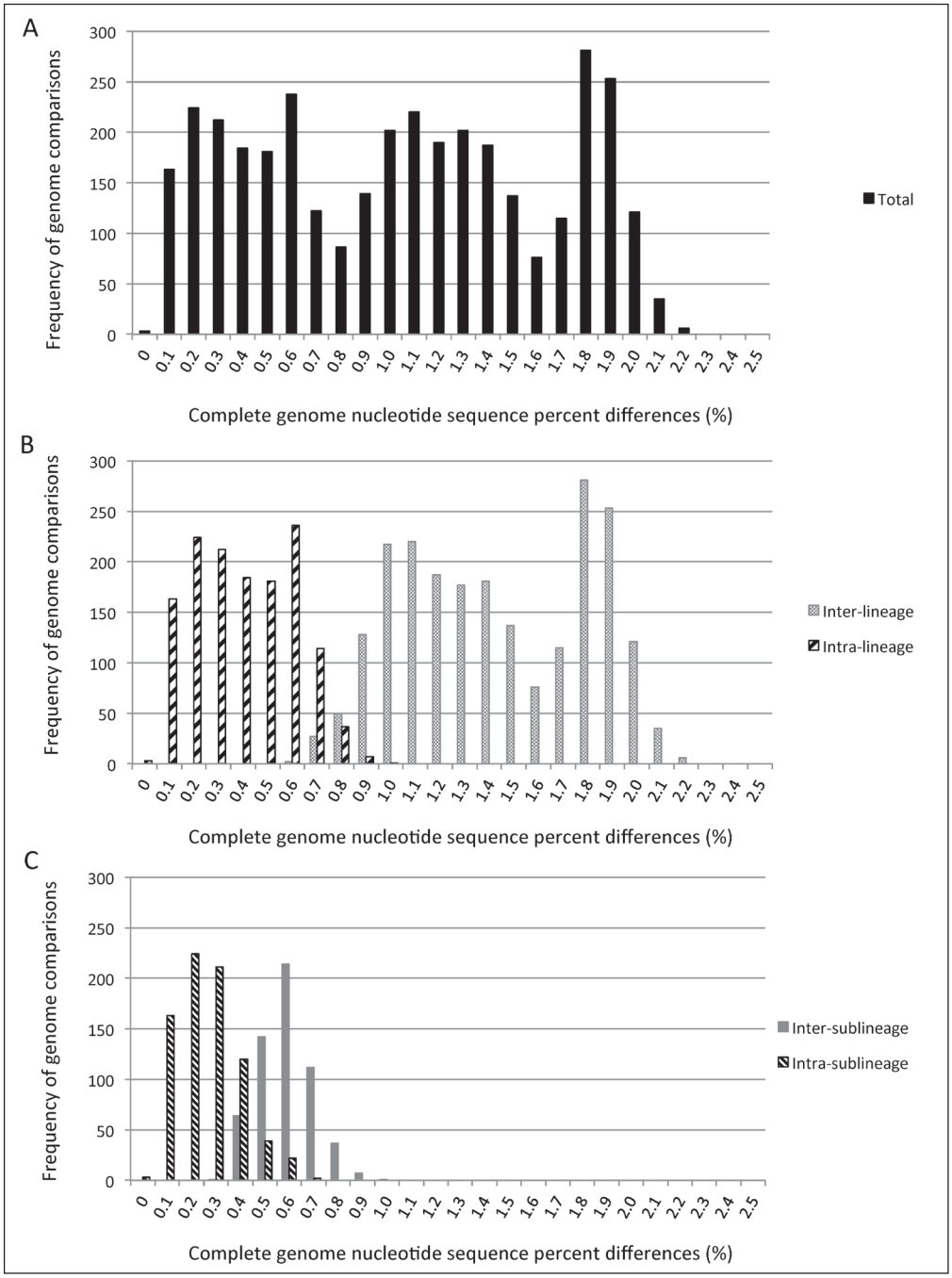

Figure 1. Distribution of pairwise differences between nucleotide sequences of alpha-9 type genomes. The genome nucleotide sequences of types 16, 31, 33, 35, 52, 58 and 67, previously reported (8, 20), were globally aligned using the program MUSCLE (21). The p-distance method in MEGA5 (22) was used to calculate the percent differences for each isolate compared to all other isolates of the same type based on a global alignment. The Y-axis represents the number of comparisons. The X-axis shows the percent nucleotide pairwise differences. (A) Comparison of each isolate to all other isolates of the same type, resulting in a total of 3577 assessments; (B) Inter- and intra-lineage pairwise differences. Inter-lineage: comparisons of isolates within different lineages of the same type (2213 comparisons). Intra-lineage: comparisons of isolates within the same lineage (1362 comparisons); (C) Inter- and intra-sublineage pairwise differences. Inter-sublineage: comparisons of isolates within different sublineages of the same lineage (578 comparisons). Intra-sublineage: comparisons of isolates within the same sublineage (784 comparisons).

Phylogenetic tree of alpha-10 species group HPV types and variant lineages

To view the relationship between the HPV members of the alpha-10 species group, a ML phylogenetic tree was constructed using representative complete genomes (Figure 4). HPV6 and HPV11 genomes form a clade and the topology indicates they shared a most recent common ancestor (MRCA).

Discussion

This is the first report describing a nomenclature for HPV6 and HPV11 variants based on complete genome analyses. HPV6 could be classified into two lineages, with the B lineage consisting of 3 sublineages (see Figure 2). HPV11 isolates were not highly variable and were classified into two sublineages (see Figure 3). The classification of variant genomes is based on a set of complete genomes, whereas the classification of HPV types is based on the L1 nucleotide sequences (4, 7). The use of full genome sequences for variant

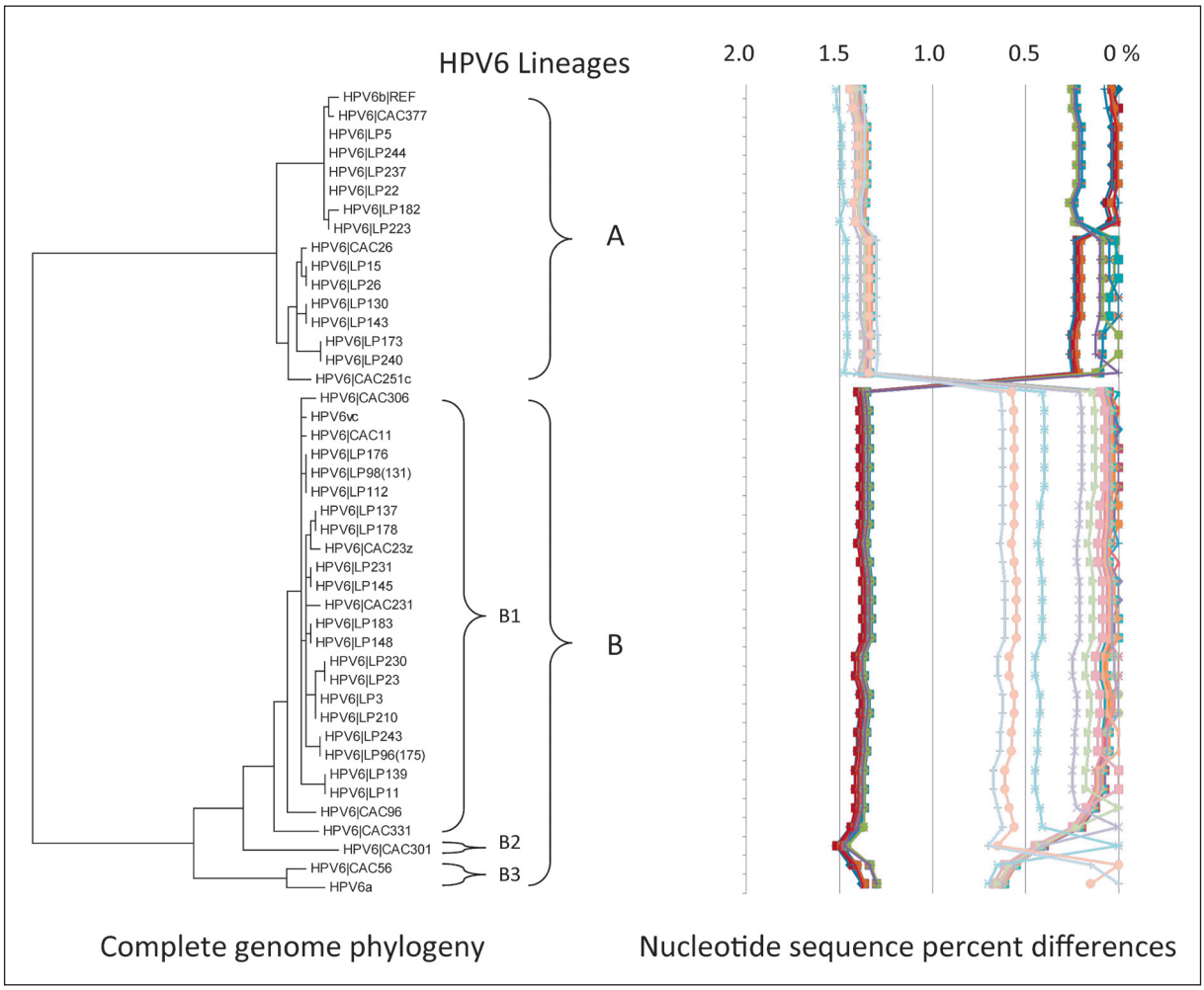

Figure 2. HPV6 variant tree topology and pairwise comparisons of individual complete genomes. A maximum likelihood (ML) tree was inferred from a global alignment of 43 complete genome nucleotide sequences of HPV6 using RAxML HPC v7.2.8 (23). Distinct variant lineages (i.e., termed A and B) and sublineages (i.e., termed B1, B2 and B3) are classified according to the topology and nucleotide sequence differences from > 1% to < 10%, and > 0.5% to < 1% ranges (4, 8). The percent nucleotide sequence differences were calculated for each isolate compared to all other isolates of the same type based on the complete genome nucleotide sequences and are shown in the panel to the right of the phylogeny. Values for each comparison of a given isolate are connected by lines and the comparison to self is indicated by the 0% difference point.

classification is derived from the fact that isolates of the same type are closely related and the full extent of the sequence heterogeneity is best summed across the whole genome; recently evolved variant genomes have changes that are not always evenly distributed throughout the genome. Nevertheless, to define distinct variant lineages, we used a nucleotide sequence difference of approximately 1.0% between two or more variants of the same type. This value was derived from empiric data on the distribution of differences between genomes of the same type from the alpha-9 species group (see Figure 1). Similarly, differences across the genome of 0.5%-1.0% were used to identify sublineages. Each variant lineage was classified and named with an alphanumeric value. The prototype or reference sequence (i.e., the cloned genome designated as the original type) is always designated variant lineage A and/or sublineage A1 (8, 28). Therefore, after the classification of variants is established based on full genome sequences, it is then possible to characterize and name isolates by the use of lineage-specific diagnostic single nucleotide polymorphisms (SNPs) or lineage-specific indels found in short sequence reads.

To facilitate the consistent numbering of nucleotide positions, we constructed a table with the key landmark nucleotide positions that can be used as a reference to name specific nucleotide variations within human alpha-PV genomes. Nucleotide positions are based on the reference sequence for each type. At some point in the past, agreement arose within the PV community that the "A" of the first ATG in the E6 ORF should be designated position "1". However, as shown in the Table 1, few of the human alpha-PVs used this criterion in naming position "1". For instance, the sequence of HPV16 defines position "1" based on an

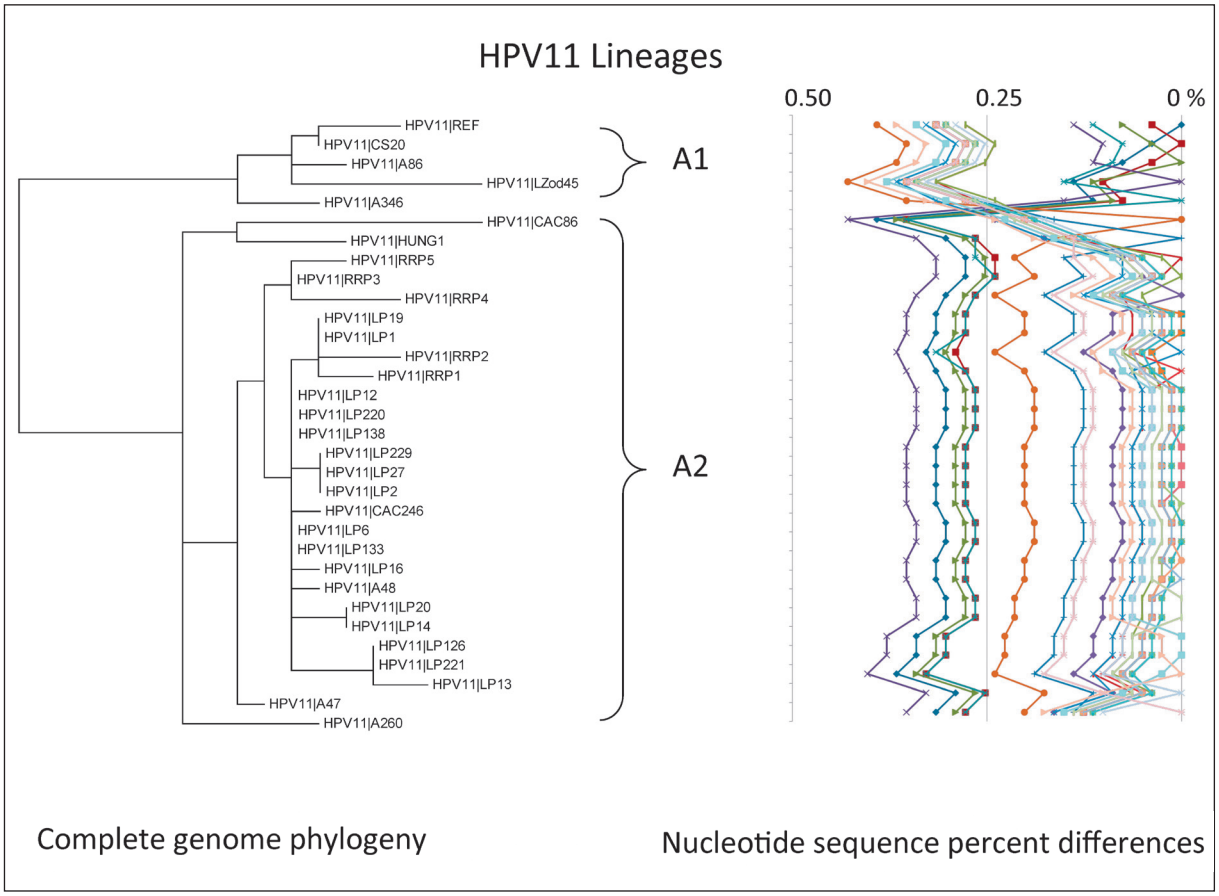

Figure 3. HPV11 variant tree topologies and pairwise comparisons of individual complete genomes. A ML phylogenetic tree was constructed from 32 HPV11 aligned complete genomes as described in Figure 2. Distinct sublineages (i.e., termed A1 and A2) were inferred from the tree topology and nucleotide sequence differences in the range of ~ 0.5%. The percent nucleotide sequence differences were calculated for each HPV11 isolate compared to all other HPV11 isolates based on the complete genome nucleotide sequences and are shown in the panel to the right of the phylogeny. Values for each comparison of a given isolate are connected by lines and the comparison to self is indicated by the 0% difference point.

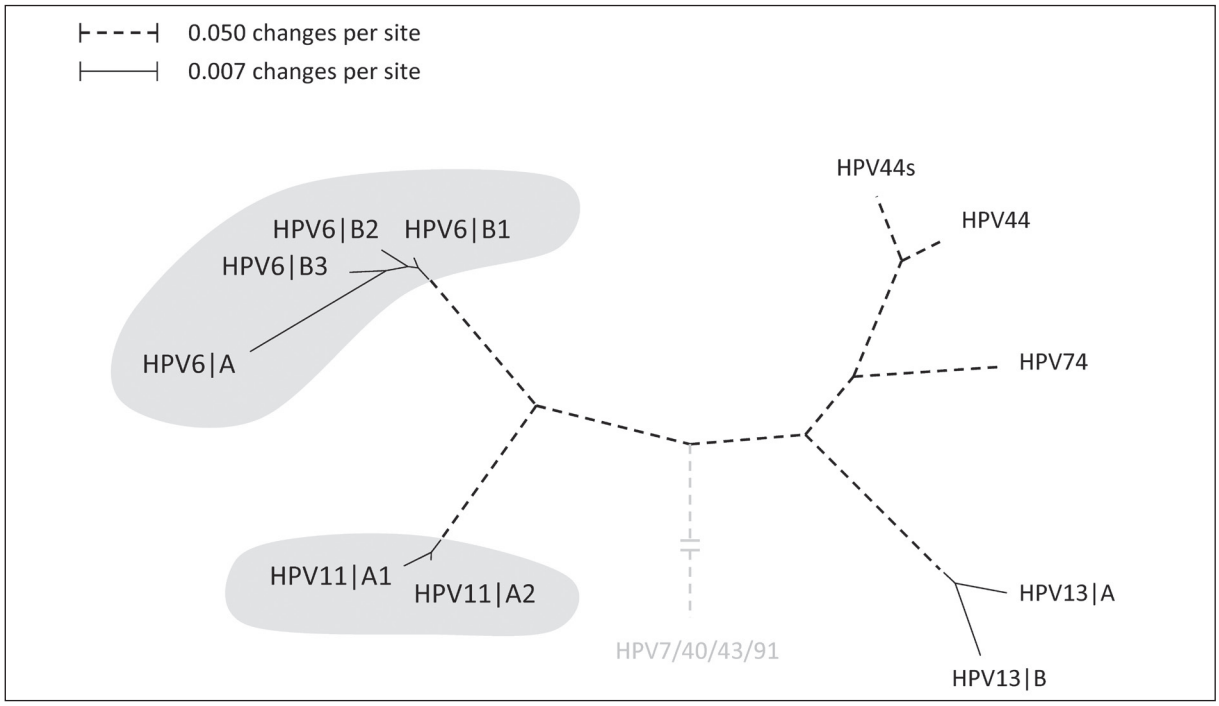

Figure 4. Alpha-10 phylgenetic tree showing representative types and variant lineages/ sublineages. A maximum likelihood tree was constructed using RAxML HPC v7.2.8 (23) inferred from the global alignment of complete genome nucleotide sequences linearized at the first ATG of the E1 ORF. Representative alpha-8 HPV types, HPV7 (NCBI accession number NC_001595), HPV40 (NC_001589), HPV43 (NC_005349) and HPV91 (NC_004085), were set as the outgroup and are shown by grey dashed lines. The shaded areas represent groupings of lineages and sublineages of HPV6 and HPV11. The length of dashed and solid lines represent distance between clades, although the number of changes is different for these two lines; the scale is indicated in the upper left corner of the figure. The GenBank accession numbers of alpha-10 HPV types are listed in the brackets following each variant: HPV6|A (X00203), HPV6|B1 (FR751337), HPV6|B2 (FR751328), HPV6|B3 (L41216), HPV11|A1 (M14119), HPV11|A2 (FN907962), HPV13|A (X62843), HPV13|B (DQ344807), HPV44 (U31788), HPV44s (U31791), HPV74 (AF436130).

alignment with the first 60 bp of HPV1a, HPV6b and BPV1 (29). Moreover, there were sequencing errors in some reference clones that have been corrected over time. For convenience, we list both the first 8 nucleotides of the reference genome and the potential E6 start codon. These 8 bp sequences can be used to search the genome to locate landmarks of interest. We recommend use of the E1 ORF ATG as a reasonably conserved site, at least, in the human alpha-PVs for multiple sequence alignments. Thus, we provide the location of the E1 ATG established from the reference sequence numbering and the 8 bp subsequent to the E1 ATG for quick identification by searching.

Variants of HPV6 form at least two deeply separated clades suggesting codivergence of host and virus as different lineages diversified from their most recent common ancestors (MRCAs). HPV11 variants are highly conserved and did not meet criterion for classification into more than one lineage. This reduced diversification probably represents a more recent divergence of HPV11 from the HPV6/11 MRCA (Figure 4). Alternatively, divergent isolates of HPV11 might exist in a remote and/or unsampled population or could have disappeared by genetic isolation and/or host demise. Another possibility is that a reduced viral population may have limited the capacity for diversification over time. In addition, previous work analyzing 62 isolates from around the world neither found a geographical association between HPV6 or HPV11 variants, nor an association with disease type (30).

A common nomenclature will allow HPV researchers to discuss the properties of HPV variant lineages without having to describe sets of nucleotide changes to define a group of HPV variants. This will be particularly useful for future studies of the alpha-10 species group of HPVs that exhibit a broad tissue tropism and have the ability to infect and cause exophytic lesions of the anogenital area, the larynx/

respiratory tract and the cervix. If we include HPV13 that causes oral focal hyperplasia, the tissue tropism of the alpha-10 species group expands to include benign infections of the oral cavity (31). To facilitate better characterization of HPV6 and HPV11 variants, different regions of the viral genome can be sequenced and the changes related back to specific lineages using the data provided in Figure 1 of Kocjan et al. (24), as they list the isolate name that corresponds to the genome sequences shown in Figure 2. To maintain a consistent nomenclature for all HPV types, we propose calling the lineage containing the original reference sequence HPV6b (32) as the "A" lineage. The other HPV6 genomes termed HPV6vc (33) and HPV6a (34) sort to sublineages "B1" and "B3", respectively.

In summary, we present a nomenclature for variants of HPV6 and HPV11. We provide a taxonomy and nomenclature of these variants that should be useful for detailed studies addressing the genetic basis of the pathogenesis of these protean HPVs that commonly cause genital warts, and/or laryngeal papillomas. The question then becomes whether different HPV6 and/or HPV11 viral lineages or specific nucleotides are associated with infection at different anatomic sites. It is likely that the magnitude of effect might be small for any lineage, thus multicenter studies will be needed to pool data and determine the phenotypes of HPV6 and HPV11 variants based on this report.

Acknowledgements

This work was supported in part by the National Cancer Institute (CA78527) (RDB), the Einstein-Montefiore Center for AIDS funded by the NIH (AI-51519) and the Einstein Cancer Research Center (P30CA013330) from the National Cancer Institute.

REFERENCES

1. Herbst LH, Lenz J, Van Doorslaer K, Chen Z, Stacy BA, Wellehan JF, Manire CA, Burk RD. Genomic characterization of two novel reptilian papillomaviruses, Chelonia mydas papillomavirus 1 and Caretta caretta papillomavirus 1. Virology 2009; 383: 131-5.
2. Terai M, DeSalle R, Burk RD. Lack of canonical E6 and E7 open reading frames in bird papillomaviruses: Fringilla coelebs papillomavirus and Psittacus erithacus timneh papillomavirus. J Virol 2002; 76: 10020-3.
3. Bennett MD, Reiss A, Stevens H, Heylen E, Van Ranst M, Wayne A, Slaven M, Mills JN, Warren KS, O'Hara AJ, Nicholls PK. The first complete papillomavirus genome characterized from a marsupial host: a novel isolate from Bettongia penicillata. J Virol 2010; 84: 5448-53.
4. Bernard HU, Burk RD, Chen Z, Van Doorslaer K, zur Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology 2010; 401: 70-9.
5. Rector A, Lemey P, Tachezy R, Mostmans S, Ghim SJ, Van Doorslaer K, Roelke M, Bush M, Montali RJ, Joslin J, Burk RD, Jenson AB, Sundberg JP, Shapiro B,Van Ranst M. Ancient papillomavirus-host co-speciation in Felidae. Genome Biol 2007; 8: R57.
6. Chen Z, Terai M, Fu L, Herrero R, DeSalle R, Burk RD. Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses. J Virol 2005; 79: 7014-23.
7. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology 2004; 324: 17-27.
8. Chen Z, Schiffman M, Herrero R, Desalle R, Anastos K, Segondy M, Sahasrabuddhe VV, Gravitt PE, Hsing AW, Burk RD. Evolution and Taxonomic Classification of Human Papillomavirus 16 (HPV16)-Related Variant Genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. PLoS One 2011; 6: e20183.
9. Scarbrough Lefebvre CD, Van Kriekinge G, Goncalves MA, de Sanjose S. Appraisal of the burden of genital warts from a healthcare and individual patient perspective. Public health 2011; 125: 464-75.
10. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011; 61: 69-90.
11. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, Meijer CJ. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 2003; 348: 518-27.

12. Li N, Franceschi S, Howell-Jones R, Snijders PJ, Clifford GM. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. Int J Cancer 2011; 128: 927-35.

13. Smith JS, Lindsay L, Hoots B, Keys J, Franceschi S, Winer R, Clifford GM. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. Int J Cancer 2007; 121: 621-32.

14. Schiffman M, Herrero R, Desalle R, Hildesheim A, Wacholder S, Rodriguez AC, Bratti MC, Sherman ME, Morales J, Guillen D, Alfaro M, Hutchinson M, Wright TC, Solomon D, Chen Z, Schussler J, Castle PE, Burk RD. The carcinogenicity of human papillomavirus types reflects viral evolution. Virology 2005; 337: 76-84.

15. Xi LF, Kiviat NB, Hildesheim A, Galloway DA, Wheeler CM, Ho J, Koutsky LA. Human papillomavirus type 16 and 18 variants: race-related distribution and persistence. J Natl Cancer Inst 2006; 98: 1045-52.

16. Sichero L, Ferreira S, Trottier H, Duarte-Franco E, Ferenczy A, Franco EL, Villa LL. High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. Int J Cancer 2007; 120: 1763-8.

17. Hildesheim A, Schiffman M, Bromley C, Wacholder S, Herrero R, Rodriguez A, Bratti MC, Sherman ME, Scarpidis U, Lin QQ, Terai M, Bromley RL, Buetow K, Apple RJ, Burk RD. Human papillomavirus type 16 variants and risk of cervical cancer. J Natl Cancer Inst 2001; 93: 315-8.

18. Ekstrom J, Bzhalava D, Svenback D, Forslund O, Dillner J. High throughput sequencing reveals diversity of Human Papillomaviruses in cutaneous lesions. International journal of cancer. Journal international du cancer 2011; 129: 2643-50.

19. Alizon S, Luciani F, Regoes RR. Epidemiological and clinical consequences of within-host evolution. Trends Microbiol 2011; 19: 24-32.

20. Smith B, Chen Z, Reimers L, van Doorslaer K, Schiffman M, Desalle R, Herrero R, Yu K, Wacholder S, Wang T, Burk RD. Sequence Imputation of HPV16 Genomes for Genetic Association Studies. PLoS One 2011; 6: e21375.

21. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004; 32: 1792-7.

22. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol Biol Evol 2011; 28: 2731-9.

23. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006; 22: 2688-90.

24. Kocjan BJ, Jelen MM, Maver PJ, Seme K, Poljak M. Pre-vaccination genomic diversity of human papillomavirus genotype 6 (HPV 6): A comparative analysis of 21 full-length genome sequences. Infect Genet Evol 2011; 11: 1805-10.

25. Maver PJ, Kocjan BJ, Seme K, Potocnik M, Gale N, Poljak M. Prevaccination genomic diversity of human papillomavirus genotype 11: a study on 63 clinical isolates and 10 full-length genome sequences. J Med Virol 2011; 83: 461-70.

26. Dartmann K, Schwarz E, Gissmann L, zur Hausen H. The nucleotide sequence and genome organization of human papilloma virus type 11. Virology 1986; 151: 124-30.

27. Wu X, Zhang C, Feng S, Liu C, Li Y, Yang Y, Gao J, Li H, Meng S, Li L, Zhang Y, Hu X, Wu X, Lin L, Li X, Wang Y. Detection of HPV types and neutralizing antibodies in Gansu province, China. J Med Virol 2009; 81: 693-702.

28. Chen Z, Desalle R, Schiffman M, Herrero R, Burk RD. Evolutionary dynamics of human papillomavirus types 18, 45 and 97 variant genomes. J Virol 2009; 83: 1443-55.

29. Seedorf K, Krammer G, Durst M, Suhai S, Rowekamp WG. Human papillomavirus type 16 DNA sequence. Virology 1985; 145: 181-5.

30. Heinzel PA, Chan SY, Ho L, O'Connor M, Balaram P, Campo MS, Fujinaga K, Kiviat N, Kuypers J, Pfister H, Steinberg BM, Tay S, Villa LL, Bernard HU. Variation of human papillomavirus type 6 (HPV-6) and HPV-11 genomes sampled throughout the world. J Clin Microbiol 1995; 33: 1746-54.

31. Cuberos V, Perez J, Lopez CJ, Castro F, Gonzalez LV, Correa LA, Sanclemente G, Gaviria A, Müller M, Sanchez GI. Molecular and serological evidence of the epidemiological association of HPV 13 with focal epithelial hyperplasia: a case-control study. J Clin Virol 2006; 37: 21-6.

32. Schwarz E, Durst M, Demankowski C, Lattermann O, Zech R, Wolfsperger E, Suhai S, zur Hausen H. DNA sequence and genome organization of genital human papillomavirus type 6b. EMBO J 1983; 2: 2341-8.

33. Kovelman R, Bilter GK, Roman A, Brown DR, Barbosa MS. Human papillomavirus type 6: classification of clinical isolates and functional analysis of E2 proteins. J Gen Virol 1999; 80 (Pt 9): 2445-51.

34. Hofmann KJ, Cook JC, Joyce JG, Brown DR, Schultz LD, George HA, Rosolowsky M, Fife KH, Jansen KU. Sequence determination of human papillomavirus type 6a and assembly of virus-like particles in Saccharomyces cerevisiae. Virology 1995; 209: 506-18.

A U T H O R S' A D D R E S S E S

Prof. Robert D. Burk, MD, Departments Pediatrics, Microbiology and Immunology, Epidemiology and Population Health, and Obstetrics, Gynecology and Woman's Health, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, New York, 10461, United States of America. Corresponding author. E-mail: robert.burk@einstein.yu.edu

Zigui Chen, PhD, Department of Pediatrics, Albert Einstein College of Medicine, same address. E-mail: zigui.chen@einstein.yu.edu

Ariana Harari, BA, Department of Microbiology and Immunology, Albert Einstein College of Medicine, same address. E-mail: ariana.harari@phd.einstien.yu.edu

Benjamin C. Smith, PhD, Department of Pediatrics, Albert Einstein College of Medicine, same address. E-mail: benjamin.smith@einstein.yu.edu

Boštjan J. Kocjan, PhD, University of Ljubljana, Institute of Microbiology and Immunology, Faculty of Medicine, Zaloška 4, 1105 Ljubljana, Slovenia. E-mail: bostjan.kocjan@mf.uni-lj.si

Polona J. Maver, MD, University of Ljubljana, Institute of Microbiology and Immunology, Faculty of Medicine, same address. E-mail: polona.maver@mf.uni-lj.si

Prof. Mario Poljak, MD, PhD, University of Ljubljana, Institute of Microbiology and Immunology, Faculty of Medicine, same address. E-mail: mario.poljak@mf.uni-lj.si