

## Discovery, characterisation and genomic variation of six novel *Gammapapillomavirus* types from penile swabs in South Africa

Alltalents T. Murahwa<sup>a,b,1</sup>, Tracy L. Meiring<sup>a,b,1</sup>, Zizipho Z.A. Mbulawa<sup>a,b,c,d</sup>,  
Anna-Lise Williamson<sup>a,b,d,\*</sup>

<sup>a</sup> Division of Medical Virology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, 7925, South Africa

<sup>b</sup> Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

<sup>c</sup> Center for HIV and STIs, National Institute for Communicable Disease, National Health Laboratory Service, Johannesburg, South Africa

<sup>d</sup> SAMRC Gynaecological Cancer Research Centre, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

### ARTICLE INFO

#### Keywords:

*Gammapapillomavirus*  
Penile

### ABSTRACT

Six novel human papillomaviruses from penile swabs were characterised. Multiple full genome clones for each novel type were generated, and complete genome sizes were: HPV211 (7253bp), HPV212 (7208bp), HPV213 (7096bp), HPV214 (7357), HPV215 (7186bp) and HPV216 (7233bp). Phylogenetically the novel papillomaviruses all clustered with *Gammapapillomaviruses*: HPV211 is most closely related to HPV168 (72% identity in the L1 nucleotide sequence) of the Gamma-8 species, HPV212 is most closely related to HPV144 (82.9%) of the Gamma-17 species, HPV213 is most closely related to HPV153 (71.8%) of the Gamma-13 species, HPV214 is most closely related to HPV103 (75.3%) of the Gamma-6 species, HPV215 and HPV216 are most closely related to HPV129 (76.8% and 79.2% respectively) of the Gamma-9 species. The novel HPV types demonstrated the classical genomic organisation of *Gammapapillomaviruses*, with seven open reading frames (ORFs) encoding five early (E1, E2, E4, E6 and E7) and two late (L1 and L2) proteins. Typical of *Gammapapillomaviruses* the novel types all lacked the E5 ORF and HPV214 also lacked the E6 ORF. HPV212 had nine unique variants, HPV213 had five and HPV215 had four variants. Conserved domains observed among the novel types are the Zinc finger Binding Domain and PDZ domains. A retinoblastoma binding domain (pRB) binding domain in E7 protein was additionally identified in HPV214. This study expands the knowledge of the rapidly growing *Gammapapillomavirus* genus.

### 1. Introduction

Human papillomaviruses (HPVs) are small non-enveloped viruses of the *Papillomaviridae* family. The genome is circular double stranded DNA of approximately 8 kb that typically contains eight genes [1,2]. PVs are classified based on sequence similarity in the gene encoding the L1 major capsid protein. HPVs of the same genus share > 60%, species 60–70% and types > 90% similarity in the L1 gene [1]. Based on this, HPVs fall into five genera; Nu, Mu, Alpha, Beta and Gamma. The genus *Gammapapillomavirus* is the most divergent and rapidly growing with 27 species and 98 officially recognised genotypes [3].

Gamma HPVs appear to have broad epithelial tissue tropism with reported detection in cutaneous, mucosal and mucocutaneous sites [4,5], including healthy skin [6], cutaneous lesions [7], gut [5], penis

[8,9], oral mucosa [10], nasal mucosa [11], anal canal [12] and cervical mucosa. These viruses have not been conclusively associated with any pathology or oncogenicity. A recent systematic review [13] did however describe significantly elevated antibody levels in squamous cell carcinomas (SCCs) cases infected with HPVs from the Gamma-1 species. While gamma HPVs remain largely unstudied at a molecular level, a recent proteomic study of HPV 197 demonstrated that the E6 and E7 proteins of this gamma HPV interact with several cellular targets including some of the important mediators of the oncogenic activities of high-risk E6 and E7 proteins [14]. The E7 protein of gamma-6 HPV 108 induces dysplasia in organotypic keratinocyte cultures [15]. The potential oncogenic activity of gamma E6 and E7 proteins requires further investigation. Interestingly HPVs in the gamma-6 species do not encode an E6 protein [16,17]. It was recently shown that gamma-6 HPVs or

\* Corresponding author. Institute of Infectious Diseases & Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Anzio Rd, Observatory, Cape Town 7925, South Africa.

E-mail address: [anna-lise.williamson@uct.ac.za](mailto:anna-lise.williamson@uct.ac.za) (A.-L. Williamson).

<sup>1</sup> These authors contributed equally.

<https://doi.org/10.1016/j.pvr.2019.02.005>

Received 30 July 2018; Received in revised form 14 February 2019; Accepted 25 February 2019

Available online 04 March 2019

2405-8521/© 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

rather E6-minus viruses have acquired a 37 amino acid protein named E10, the open reading frame (ORF) encoding this protein is upstream of the E7 start codon [18,19]. However, the acquisition of E10 does not likely compensate for all E6 functions.

In the present study we report the genomic characterisation and phylogenetic evaluation of six novel *Gammmapapillomaviruses* types: HPV211, HPV212, HPV213, HPV214, HPV215 and HPV216. These HPVs were previously identified from a study [20] that was done on 218 penile samples (104 HIV negative and 114 HIV positive) using high throughput sequencing (Roche 454) of amplicons obtained using FAP59/64 primers which were designed to detect “cutaneous” or Beta- and *Gammmapapillomaviruses* [21]. In the study fifteen putative novel HPV types, including HPV211 (CT02, KY063000), HPV212 (CT03, KY063001), HPV213 (CT04, KY063002), HPV214 (CT06, KY063004), HPV215 (CT07, KY063005), HPV216 (CT12, KY063010), were identified from the short HPV L1 FAP fragments with a prevalence varying between 0.5% and 4.1% of men sampled [20]. We further examined variation of the novel types in clinical specimens from which they were identified.

## 2. Methods and materials

### 2.1. Ethics statement

Ethical approval for the study was granted by the Health Research Ethics Committee of the University of Cape Town, Faculty of Health Sciences (HREC reference: 231/2015 and 258/2006). Written consent was obtained from all the study participants.

### 2.2. Penile samples

Samples were obtained from African black men who were participating in a heterosexual couples study on the natural history of HPV infection in South Africa [22]. Penile swab samples were obtained by dry swabbing of the penile shaft, glans and foreskin, if present, using a Digene swab and stored at  $-80^{\circ}\text{C}$  in specimen transport medium (STM, Qiagen) HPV211 originated from a 29 year old HIV negative, HPV212 from a 42 year old HIV positive, HPV213 from a 28 year old HIV negative, HPV214 from a 24 year old HIV negative, HPV215 from a 41 year old HIV negative and HPV216 from a 45 year old HIV positive individual.

### 2.3. Nucleic acid isolation and amplification

DNA was extracted using the MagNA Pure Compact Nucleic Acid Isolation kit (Roche, USA) and circular genomes enriched using the Illustra TempliPhi 100 Amplification kit (GE Healthcare, Amersham, UK), according to the manufacturer's instructions.

### 2.4. Complete genome amplification

The complete genomes were amplified by PCR using back to back primers (Table 1) and the KAPA LongRange HotStart PCR kit (KAPA Biosystems, USA). Primers were designed based on the L1 FAP sequences from Meiring and co-workers. 2017 [20].

Each reaction had 1X KAPA LongRange Buffer, 1.75 mM  $\text{MgCl}_2$ , 0.3 mM of each dNTP, 0.675 U of the KAPA LongRange DNA polymerase, 0.5  $\mu\text{M}$  of each primer and nuclease free water. The PCR conditions were as follows:  $94^{\circ}\text{C}$  for 3 min followed by 12 cycles of 25 s at  $94^{\circ}\text{C}$ ,  $68^{\circ}\text{C}$  (reduced by  $2^{\circ}\text{C}$  every 2 cycles) for 15 s, and 7 min of extension at  $68^{\circ}\text{C}$ , followed by 20 cycles of 25 s at  $94^{\circ}\text{C}$ , 56  $^{\circ}\text{C}$  for 15 s and 7 min at  $68^{\circ}\text{C}$ . The final extension was performed at  $72^{\circ}\text{C}$  for 10 min.

### 2.5. Cloning and sequencing

PCR products were gel purified using the MinElute Gel Extraction

**Table 1**

Novel HPV L1 back to back type specific primer sequences.

HPV type	Primer Sequences
HPV211	forward-GTTACGGGGAATTCAGATAGGTAGAGGTGG reverse- TTCCAAACCAGTCTTTCATGGTCAGAATTG
HPV212	forward- ATAGAAATAGGTAGAGGTGGGCCCTTAGG reverse-AGCTCGTAATTTCCAAACTAAGCGTTCG
HPV213	forward-TGGCAGTTACGGGGTGTGTAGGTAGAC reverse- CACAAGACGTTCTCGTTGGGATCATAAC
HPV214	forward-GCCTTTGGGTATTGGGTCTACTGTGCAC reverse-CCACCCCGTGCAATATCAATGCCACGTAAC
HPV215	forward- ATGGTTTGCAAATTGACAGAGGTGGTCC reverse- ATAATTTCCACACCAGCCTTTCATGTTGTG
HPV216	forward- CAACAGGTCATCCATTATTGATCGCTTAC reverse- TTCCTATCCCTAGGGACCACCTCTATC

Kit (Qiagen, Germany) and cloned into the pGEM-T Easy Vector System (Promega, USA). Ten clones of each of the HPV types were selected for sequencing. Illumina sequencing libraries were prepared using the KAPA HyperPlus kit (Roche, USA) according to the manufacturer's instructions using TruSeq Dual Index Illumina adaptors (Illumina, USA). Illumina MiSeq 300 bp paired end sequencing was carried out by Macrogen Inc. (Seoul, South Korea).

### 2.6. Illumina data quality control and assembly

A total of 32,409,876 reads, with a median of 578,992 (range: 352,306–784,098) reads per sample, were included in the analysis. Illumina sequence reads were processed in CLC Genomic Workbench (Qiagen, Germany). Reads were trimmed using a quality score limit of 0.05 and any reads with more than 2 ambiguities discarded. Reads were assembled using the *de novo* assembly function with default parameters and vector sequences trimmed in CLC Genomics Workbench (Qiagen, Germany). Coverage plots were also done using CLC Genomics (see Supplementary Fig. S2).

### 2.7. Nucleotide accession numbers

The novel HPV genome sequences were deposited in Genbank under the following accession numbers: HPV211 MF509816, HPV212 MF509817, HPV213 MF509818, HPV214 MF509819, HPV215 MF509820, and HPV216 MF509821. The sequence reads were also deposited in the Sequence Read Archive (SRA) under the following accession numbers: SRX5369055 to SRX5369122.

### 2.8. Genomic characterisation

ORF prediction and multiple sequence alignments were done in CLC Genomics Workbench (Qiagen, Germany). The first position of the complete genomes of HPV211, HPV212, HPV213, HPV215 and HPV216 were set at the first ATG of the E6 ORF. For HPV214 the first position was set at the first ATG of the E7 ORF because it lacks the E6 ORF. E1<sup>E4</sup> and E2<sup>E8</sup> spliced gene products were predicted using manual inspection of splice sites predicted by NNSplice 0.9 [23] filtered using the set of criteria defined by van Doorslaer and co-workers [19].

Conserved domains were predicted by directly searching the expected conserved protein or nucleotide sequences in multiple sequence alignments with homologous viral regions or proteins from the closest related HPV types.

### 2.9. Phylogenetic analysis

For the phylogenetic analysis L1 sequences from HPV211-HPV216 and the Papillomavirus episteme (PAVE) database [24] were aligned with MUSCLE [25]. A maximum likelihood tree was generated with PhyML 3.0 [26] using the GTR + I + R substitution model, as determined

by jmodeltest [27]. The approximate likelihood ratio test (aLRT) was used to estimate branch support [28]. The tree was visualised in iTOL (<http://itol.embl.de/upload.cgi>) [29].

In order to select the most suitable clone to submit to the International HPV reference laboratory (<https://ki.se/en/labmed/international-hpv-reference-center>) for purposes of nomenclature, multiple whole genome sequence alignments of the clones of each novel type were done and the clone with the highest percentage pairwise identity to the consensus sequence was treated as the prototype reference sequence. The prototype reference clone was sent to the International HPV Reference Centre who assigned HPV numbers on verification of the sequence (HPV211 to HPV216).

### 2.10. Nucleotide and amino acid variation

Nucleotide and amino acid variations in novel HPV types that showed differences in the clones generated were determined from multiple sequences alignments of the clones of each novel type. Percentage nucleotide and amino acid substitutions were also determined for each ORF. The selection pressures acting on the coding sequences of the novel HPVs were estimated by calculating codon-specific non-synonymous ( $d_N$ ) and synonymous ( $d_S$ ) substitution rates using the program SNAP version 2.1.1 [30] (<http://www.hiv.lanl.gov>). This program uses the unweighted pathway method of Nei-Gojobori [31] and the Jukes-Cantor model [30]. The  $d_N/d_S$  ratios were calculated, with a ratio of 1 indicating neutral selection, > one diversifying positive selection and < one negative or purifying selection [32].

## 3. Results

### 3.1. Genomic organisation and phylogeny

In this study the full genomes of six novel HPVs were cloned into pGEM-T vector and sequenced using Illumina MiSeq 300 bp paired end sequencing. The assembled sequences of the viral genomes revealed a total size of 7253 bp for HPV211, 7208 bp for HPV212, 7096 bp for HPV213, 7357 for HPV214, 7186 for HPV215 and 7233 for HPV216 (Table 2).

Pairwise comparison of all the novel types showed less than 60% identity to each other meaning they are members of the same genus but different species with the exception of HPV215 and HPV216 (72%) which belong to the same species, Gamma-9 [1]. HPV211 was most closely related to HPV168 with a pairwise identity of 72% based on the L1 nucleotide sequence. HPV212 was most closely related to HPV144 with a pairwise identity of 82.9%, HPV213 was most closely related to HPV153 with a pairwise identity of 71.8%, HPV214 was most closely related to HPV103 with a pairwise identity of 75.3%, HPV215 was most closely related to HPV129 with a pairwise identity of 76.8% and lastly HPV216 was most closely related to HPV129 with a pairwise identity of

79.2% (Table 3).

HPV211 clustered in the Gamma-8, HPV212 in the Gamma-17, HPV213 in the Gamma-13 and HPV214 in the Gamma-6 species as shown in the maximum likelihood tree in Fig. 1.

The novel HPV types demonstrated the classical genomic organisation of *Gammapapillomaviruses*, with seven ORFs encoding five early (E1, E2, E4, E6 and E7) and two late (L1 and L2) proteins. Typical of *Gammapapillomaviruses* the novel types all lacked the E5 ORF. HPV214 additionally lacked the ORF encoding the E6 protein but a putative ORF called E10, upstream of the E7 ORF (Supplementary Fig. S1a) was identified [18,19]. HPV212 did not have an E4 start codon as has been reported elsewhere [33]. For all the putative ORFs, nucleotide and amino acid similarities to the closest recognised HPV types are shown in Table 3, and all were closely related to members of the *Gammapapillomavirus* genus.

### 3.2. Conserved domains

The putative E6 protein of all the novel types, except for HPV214 that did not have the E6 ORF, contained two zinc-finger domains (CxxC(x)<sub>29</sub>CxxC) separated by 36 amino acids and the E7 contained one zinc finger domain (Table 4). Putative PDZ binding domains were also identified in the E6 N-terminal of all the novel types except for HPV212 that did not show the classical x(T/S)x(L/V) PDZ motif (Table 4). The LxCxE motif [34,35] essential for binding of retinoblastoma protein (pRB) was present only in the E7 protein of HPV214 (Table 4). The E1 ORF of all the novel types contained the ATP binding sites, nuclear localisation like signal (NLS-like) sequences and the nuclear export like sites, while the E2 proteins all lacked the leucine zipper domain (L(x)<sub>6</sub>L(x)<sub>6</sub>L(x)<sub>6</sub>L(x)<sub>6</sub>L), but had the NLS-like sequence and the DNA recognition helix. The DNA recognition motif in the E2 protein of HPV214 has a slightly different motif from GxxNxLKCxRxR(x)<sub>8</sub> to the GxxNxTKCxRxR(x)<sub>8</sub>, with an Leucine to Threonine modification and the functionality of this motif can only be investigated by functional assays.

Both the L1 and L2 proteins of the novel types contained NLS-like sequences furthermore the L2 protein also contained the transmembrane binding domain and the furin cleavage site. We also identified early polyadenylation sites (-AATAAA-) at the N-terminal end of the L2 protein of all the novel HPV types. The late polyadenylation sites were located in the LCR region but at differing positions in each novel type. Typical of Gamma-HPVs the LCR of all the novel HPV types contained the palindromic E2 binding sites ACC-(N)6-GGT as shown in Table 4.

### 3.3. Novel HPVs nucleotide and amino acid variations

All the clones of HPV211, HPV214 and HPV216 were identical and showed 100% pairwise identity. The clones of HPV213, HPV215 and of HPV212 had several differences. Analysis of mismatches between the 9 genomic clones of HPV212 showed a total of 67 positions that varied

**Table 2**

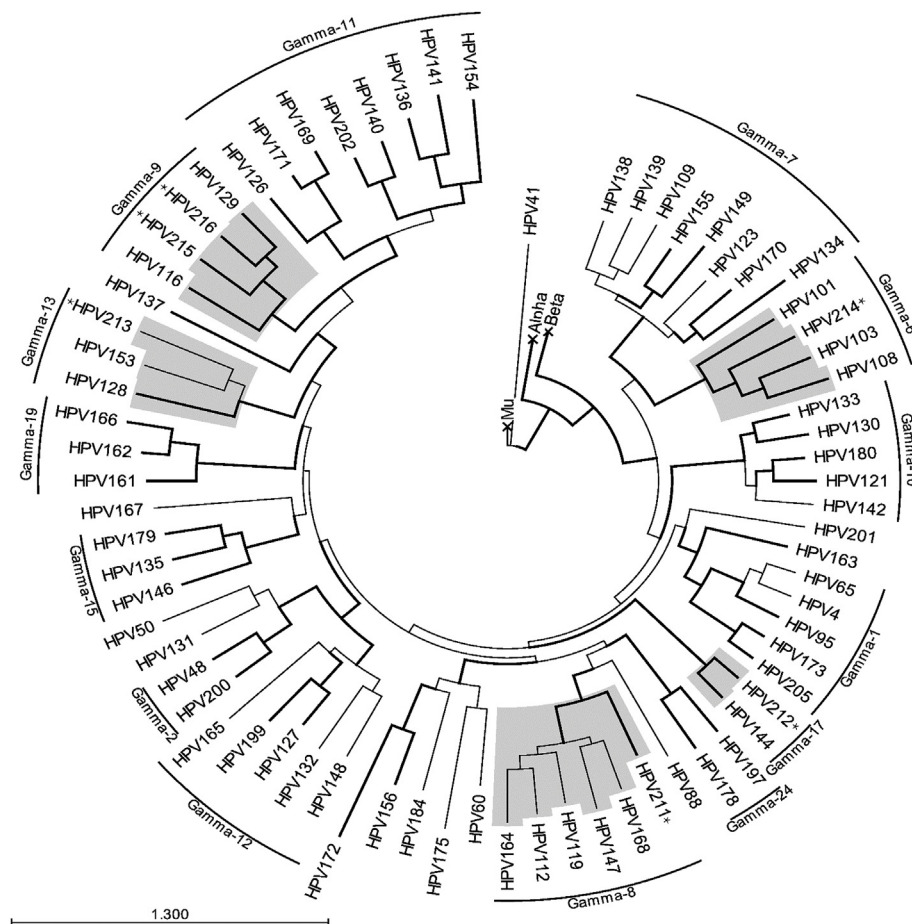
Genome lengths of novel HPV types and ORF positions on the genome and sizes of proteins.

	HPV211	HPV212	HPV213	HPV214	HPV215	HPV216
Genome length (bp)	7253	7208	7096	7357	7186	7233
% GC	37.6	38.2	39.8	41.6	38.6	37.4
<b>ORF position (Protein size in amino acids)</b>						
E6	1-420 (140)	1-447 (149)	1-435 (151)	–	1-438 (146)	1-459 (153)
E7	417-707 (97)	444-743 (100)	422-715 (98)	1-303 (101)	440-733 (98)	462-755 (98)
E1	694-2508 (605)	727-2526 (600)	702-2513 (604)	284-2188 (635)	717-2528 (604)	739-2550 (604)
E2	2429-3619 (397)	2465-3628 (388)	2449-3624 (392)	2130-3296 (389)	2473-3624 (384)	2495-3667 (391)
E4 <sup>a</sup>	2940-3380 (147)	< 3036–3386(117)	2921-3397 (159)	2704-3057 (118)	2900-3391 (164)	2922-3434 (171)
L2	3622-5220 (533)	3628-5154 (509)	3629-5122 (498)	3360-4973 (538)	3626-5137 (504)	3669-5186 (506)
L1	5231-6781 (517)	5166-6719 (518)	5133-6659 (509)	4984-6519 (512)	5146-6696 (517)	5195-6748 (518)

<sup>a</sup> start codon not determined for HPV212.

**Table 3**  
Nucleotide (amino acid) percentage identity to closest HPV type.

	HPV211	HPV212	HPV213	HPV214	HPV215	HPV216
HPV type most closely related (% L1 identity)	HPV168 (72.0)	HPV144 (82.9)	HPV153 (71.8)	HPV103 (75.3)	HPV129 (76.8)	HPV129 (79.2)
E6	66 (51)	87 (89)	72 (56)	–	73 (69)	84 (83)
E7	70 (55)	88 (87)	73 (61)	74 (69)	77 (74)	80 (76)
E1	69 (61)	88 (92)	75 (71)	72 (65)	79 (77)	85 (88)
E2	67 (55)	90 (90)	72 (56)	70 (62)	75 (70)	82 (76)
E4	69 (44)	89 (81)	79 (40)	72 (63)	74 (60)	83 (75)
L2	65 (53)	80 (85)	70 (52)	71 (62)	69 (65)	77 (75)
L1	72 (71)	83 (92)	70 (68)	75 (75)	77 (80)	79 (87)



**Fig. 1.** Maximum likelihood tree of the novel HPV types and related types generated by Muscle alignment [25] of L1 nucleotide sequences in PHYML [26] with GTR substitution. Branch support > 80% is indicated by thick branch lines. Asterisks indicate the novel HPVs. The Gamma-HPV species are named according to Bzhalava and co-workers [74] and the species including the novel types are highlighted in grey.

along the 7208 bp genome (Fig. 2A). All the HPV212 clones were unique. Analysis of mismatches between the 10 genomic clones of HPV213 showed a total of 51 positions that varied along the 7096 bp genome (Fig. 2B). HPV213 had 5 unique clones. Analysis of mismatches between the 6 genomic clones of HPV215 showed a total of 50 positions that varied along the 7186 bp genome (Fig. 2C). HPV215 had 3 identical and 3 different clones see also Fig. 2C. All the clones of each type showed between 99.5% and 100% pairwise identity to each other (Fig. 3).

Nucleotide changes that resulted in amino acid changes in the different ORFs are shown in Fig. 4. There were no nonsynonymous changes in conserved functional domains except for two HPV213 clones that had a change from Cysteine to Phenylalanine at the last position of the E7 zinc finger binding domain. In summary the percentage amino acid substitutions for HPV212 clones ranged from 0.98% in the L2 ORF to 2.67% in the E1 ORF, for HPV213 clones from 0% in the E7 ORF to 1.18% in the L1, and lastly in novel HPV215 from 0.56% in the E4 to 1.79% in the L2. In Table 5 below it is shown that all the ORFs of the

novel HPV types had  $d_N/d_S$  ratios less than 1, which is indicative of negative selection pressure or purifying selection.

#### 4. Discussion

Pertaining to the phylogenetic tree presented in Fig. 1, the novel HPV types are placed within their different species of the gamma PVs with high branch support values. HPV211 clustered in the Gamma-8 species alongside five other members. HPV212 becomes the second member of Gamma-17 species after the sole member HPV144 isolated from an oral rinse [36]. HPV213 becomes the third member of the Gamma-13 species after HPV153 isolated from a condyloma [37] and HPV128 isolated from a skin wart [38]. HPV215 and HPV216 become third and fourth members of the Gamma-9 species after HPV129 isolated from a skin wart and HPV116 isolated from a rectal swab [39]. HPV214 becomes the fourth member of Gamma-6 species after HPV101, HPV103 isolated from cervico-vaginal cells [17] and HPV108 also isolated from a cervical lesion [15]. HPV214 was also identified as



**Table 4**  
Presence and frequency of potential conserved domains in the novel HPV types.

Regions	Domain	Motif	HPV211	HPV212	HPV213	HPV214	HPV215	HPV216
E6	Zinc finger binding domain	CxxC(x) <sub>29</sub> CxxC	2	2	2	NA	2	2
	PDZ binding domain	x(T/S)x(L/V)	1	–	1	NA	1	1
E7	Zinc finger binding domain	CxxC(x) <sub>29</sub> CxxC	1	1	1	1	1	1
	pRB binding domain	LxCxE	–	–	–	1	–	–
E1	ATP binding site	G(x) <sub>4</sub> GK(T/S)	1	1	1	1	1	1
	Bipartite nuclear localisation signal	KRK and KRRL	1	1	1	1	1	1
	Nuclear export domain putative	(L/I)(x) <sub>2,3</sub> (L/I)xx(L/I/V)x(L/I/V)	1	–	1	1	1	1
E2	Leucine zipper domain	L(x) <sub>6</sub> L(x) <sub>6</sub> L(x) <sub>6</sub> L	–	–	–	–	–	–
	DNA recognition helix	GxxNxLKCxRxR(x) <sub>8</sub>	1	1	1	1 <sup>a</sup>	1	1
	Nuclear localisation domain	RKRxR/KRRR/KRXR	1	1	1	1	1	1
L1	Nuclear localisation like domain	K(K/R)R(K/R)	1	1	1	1	1	1
L2	Nuclear localisation like domain	(K/R) <sub>3</sub> R(K/R)	1	1	1	1	1	1
	Transmembrane binding domain	G(x) <sub>3</sub> G(x) <sub>3</sub> G	1	1	1	1	1	1
LCR	Furin cleavage site	Rx(K/R)R	1	1	1	1	1	1
	Early polyadenylation site	AAT(A) <sub>3</sub>	1	–	1	1	1	1
	E2 binding sites	ACC(N) <sub>6</sub> GGT	3	3	1	3	3	2
	TATA binding box	TAT(A) <sub>3</sub>	1	1	1	1	1	1
	Late polyadenylation site	AAT(A) <sub>3</sub>	1	1	1	1	1	1

N represents any nucleotide, and x represents any amino acid.

LCR: long control region. Retinoblastoma binding protein (pRB).

<sup>a</sup> The DNA recognition motif of HPV214 E2 has a L to T modification GxxNxTKCxRxR(x)<sub>8</sub>.

**A**

ORF	E6	E7	E1	E2	E4	L2	L1	LCR
	2 3	4 4 4 5	1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2	2 2 2 2 2 3 3 3 3 3 3	3 3 4 4 4 5 5 5 5 5	5 5 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6	6 6 6 6 6 7 7 7 7 9 0 1	6 6 6 6 7 7
Variants	2 6 9 2 8	6 7 9 2	0 2 3 6 9 3 6 9 7 0 3 1 5 4 6 3 3 5	1 4 3 6 6 1 2 8 6 6	3 3 7 7 0 2 2 3	0 1 8 7 8 9 4 4 4 3 8 1 3 3 1	3 6 9 3 6 8	7 9 8 0 7 2
HPV212-1	A A T T A	T A A G	T A A A A A A A A A A A C C T G A T T	G G T T T A A C T T	A T A A T G A A	T A G T T T T G T A T A G T T	T C T T T T	
HPV212-2	...	...	... T T ...	... C . C G . T C C	... . . . . A . G	C G . . . . . A . . . . .	... T ...	
HPV212-3	G . . . G	. G . .	... G G . G . . T C . . . .	... C . . . G T . .	G . . . . C . . .	C . . . . . C . . . . .	C T . . . . C	
HPV212-4	...	...	... G . . . . . T . . C .	... C . . . . T . .	... . . . . G .	C . . C C . C . . . . C .	. T . . . .	
HPV212-5	...	. . . G .	... G . C . T . . . . .	... C . . . . T . .	... . . . . .	C . . . . . G . . . . .	. T C C . .	
HPV212-6	. . . C .	...	... G . . . . . T . . G . C	. A C . . . . T . .	. . . G G . . . .	C . . . . . . . . . .	. T . . C .	
HPV212-7	. G . . .	...	. . . G G . . . . . T . A . . .	. . . C C . . T . .	... . . . . .	C . . . . C . . . . .	. T . . . .	
HPV212-8	...	. . . C . . A	... . . . . . T . . . . .	. . . A . C . . . T . .	. . . C . . . . .	C . . . . . C G . . . .	. T . . . .	
HPV212-9	. . C . .	...	. . . C . . . G . . . . . T . . . . .	. . . C . . . T . .	... . . . . .	C . A . . . . A . . C . . . .	. T . . . .	

**B**

ORF	E6	E7	E1	E2	E4	L2	L1	LCR
	1 4	6	7 8 8 9 9 9 2 4 5 5 6 6 6 6 7 0 2	6 6 6 8 8 1 2 4 5 5	1 1 5 7 7 8 0	3 3 3 5 5 6 7 7 1 4 5	7 8	
Variants	1 2 0	7	2 5 6 2 5 6 0 2 2 8 1 1 1 9 3 2 5	2 5 7 3 6 6 7 5 3 4	6 0 7 2 7 8 2	5 6 7 1 7 8 0 4 9 1 4	8 4	
HPV213-7/8	A T G	G	T A G A A A A C A G T C A T T T T	C A T T T A A T G A	G A T T T G A	T G A T T T T A T A C	A T	
HPV213-1/4	. . A	. .	. . G A G . . . . G . . . . . C .	. . . . . C . .	A . . A . A .	. A . . . . . G . . . .	. .	
HPV213-2/3	. . A	. .	. . C . A . . G . . . A . . . C . . .	. . . . . G . . . . .	A . . A . A .	. A . C . . . G A . T . C	. .	
HPV213-5/6	. . A	A	. . . A . . . . . C . . . . .	. . . G . . C . . . . .	A G C A C A	C A . . C C . G . . . .	. .	
HPV213-9/10	T C A	. .	. . A . G . G G . . C T T . . . C	A . C C . . G . T C	A . . A . A G	. A G . . . G G . G . G .	G .	

**C**

ORF	E6	E7	E1	E2	L2	L1	LCR
	1 1	6 7 7	8 9 1 1 3 4 4 5 7 0 4 4	5 7 7 8 2 4	6 6 7 8 9 9 0 0 1 1 4 6 7 7 7	2 3 7 7 7 8 9 2	7
Variants	5 8 0 9	4 0 2	1 0 1 4 5 6 8 7 9 6 3 7	5 0 1 5 1 1	3 6 7 5 4 4 8 9 8 9 9 3 1 3 3	9 4 2 4 8 1 0 0 4	4
HPV215-5	T C A T	C A A	A T G T G A A T T G A A	A T A A C A	A T T T A T A T G C C A T G C	C T G A T C T T	T
HPV215-1	C T T C	T . G	G . A C A G G C . A . . . . .	. . . . .	. . . . . G C . C A . . . . A T	T C . . . . .	. .
HPV215-2/3/4	. T T C	T . .	. . . . A . . . . A . . . . .	. . A G G . . . .	. . . . . G C G . A . . . . A T	T . A . C T C . .	. .
HPV215-6	. T T C	T G .	C C . . A . . C A G G	G . . . T G	G G C C G C . . A T T G C A T	T . . G . . . C	C

**Fig. 2.** Nucleotide sequence variations across the complete genomes of the different clones of HPV212 (with clone 1 as reference), HPV213 (with identical clones 7 and 8 as reference clones) and HPV215 (with clone 5 as the reference). Similarities are represented by dots and differences by the nucleotide changes. LCR-upstream regulatory region.

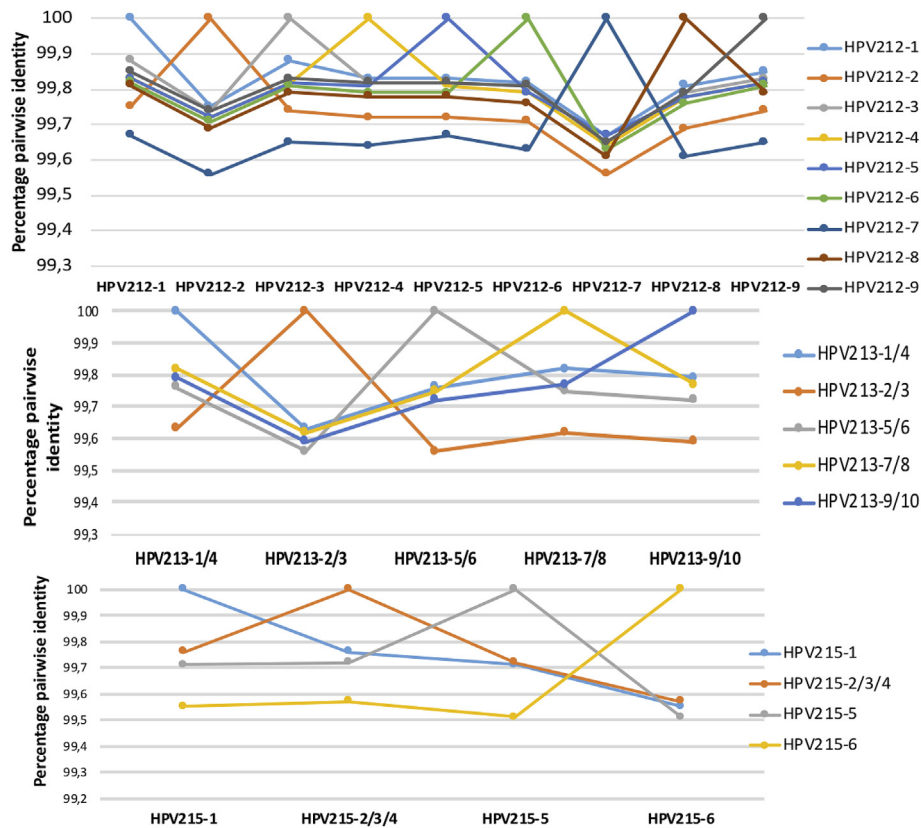


Fig. 3. Percentage pairwise identities of the different full genome clones of HPV212 (A), HPV213 (B) and HPV215 (C). Values of each pairwise comparison of the clones are connected by lines (coloured differently for each clone) and comparison to self is indicated by the 100% pairwise identity point.

ORF	E6	E7	E1	E2	L2	L1
Reference	9 23 128	7 29	61 67 137 189 204 212 255 316 327 336 474 479 530 538 544 559	60 89 133 233 334 366	103 168 448 459 465	15 245 302 310 327
HPV212-2	.	.	.	. A . A P T	. . . . T	P . . . .
HPV212-3	V . R	.	. . . . R V . C . A . . . .	. A . . . .	V . . L .	P . . . .
HPV212-4	.	.	. . . . G . . . . . L . . . .	. A . . . .	. . . . .	P . . . .
HPV212-5	.	.	. . . . . G . H . . . . .	A . . . . .	. . . . .	P . S . .
HPV212-6	.	.	. . . . . . . . . . R . L .	I A . . . . .	. R . . .	P . . P .
HPV212-7	. A . . . .	. . . . .	G D . . . . . . . . . .	. A H . . . . .	. . . . .	P . . . .
HPV212-8	.	T . . . . .	. . . . . . . . . . .	. A . . . . .	. . . . .	P . . . .
HPV212-9	.	.	P . . S . . . . . . . . . .	. A . . . . .	. G . . . .	P S . . H

ORF	E6	E7	E1	E2	E4	L2	L1
Reference	134	86	56 84 293 296 305 518	70 141 240 337 363 366	149	181 465	75 78 81 129 149 185 190 206 471
HPV213-8	G	C	G K A C A I	N Y S L F K	H	T S	H R D S M C L E T
HPV213-1/4	R	.	. . . . .	. . . . P . . . .	.	.	. . . . . K . . . . G .
HPV213-2/3	R	.	E . . Y . . . . .	. . . . G . . . .	.	.	. . . . . K . P . . . G I
HPV213-5/6	R	Y	. . . . .	S H . . . . .	.	A	L K . . T R . G .
HPV213-9/10	R	.	E R V . V T . . . . .	. . L N L N	R	. G	. K G . . . V G .

ORF	E6	E7	E1	E2	E4	L2	L1
Reference	20 29 36	97	34 212 249 258 287 361 573 587	2 27 82 248	120	2 12 25 106 108 154 158 190 369	50 199 215 223 236
HPV215-1	. L F	.	. T . . . A E G	. . . . .	.	. . . A A . A . T	Y . . . .
HPV215-2/3/4	. L F	.	. T . . . . .	. G E . . . .	.	. . . A A R . . T	Y . S S S
HPV215-6	S L F	D	G T R R P . . . .	D G . M	C	R G S A A . . I T	Y G . . .

Fig. 4. Amino acid variations in the predicted proteins of the HPV212 clones (with clone 1 as reference), HPV213 clones (with identical clones 7 and 8 as reference clones) and HPV215 clones (with clone 5 as the reference).

**Table 5**

Comparison of nucleotide and amino acid sequence variability and synonymous to nonsynonymous amino acid changes within HPV212, HPV213 and HPV215 proteins.

ORF	Number of nucleotides	Number (%) of variable nucleotide positions	Number of amino acids	Number (%) of nonsynonymous changes	Number of synonymous Changes	$d_N/d_S$ ratio
<b>HPV212</b>						
E6	447	5 (1.12%)	149	3 (2.01%)	2 (1.34%)	0.47
E7	300	4 (1.33%)	100	2 (2.00%)	2 (2.00%)	0.00
E1	1800	19 (1.06%)	600	16 (2.67%)	3 (1.50%)	0.67
E2	1164	9 (0.77%)	388	6 (1.55%)	3 (0.77%)	0.27
E4	378	3 (0.79%)	126	3 (2.38%)	0 (0%)	0.00
L2	1527	9 (0.59%)	509	5 (0.98%)	4 (0.79%)	0.45
L1	1554	15 (0.97%)	518	6 (1.16%)	8 (1.54%)	0.45
LCR	489	6 (1.23%)	–	–	–	–
<b>Total</b>	<b>7208</b>	<b>67 (0.92%)</b>	<b>2390</b>	<b>41 (1.72%)</b>	<b>22 (0.92%)</b>	
<b>HPV213</b>						
E6	454	3 (0.66%)	151	1 (0.66%)	2 (1.32%)	0.14
E7	294	1 (0.34%)	98	1 (1.02%)	0(0%)	0.00
E1	1812	17 (0.94%)	604	9 (1.49%)	8(1.33%)	0.16
E2	1176	10 (0.85%)	392	6 (1.53%)	4 (1.02%)	0.20
E4	564	3 (0.53%)	188	1 (0.53%)	4 (2.13%)	0.54
L2	1494	7 (0.47%)	498	2 (0.40%)	5 (1.00%)	0.12
L1	1527	11 (0.72%)	509	9 (1.77%)	2 (0.39%)	0.55
LCR	437	2 (0.46%)	–	–	–	–
<b>Total</b>	<b>7096</b>	<b>51 (0.72%)</b>	<b>2440</b>	<b>29 (1.19)</b>	<b>25(1.02)</b>	
<b>HPV215</b>						
E6	438	4 (0.91%)	146	3 (2.05%)	1 (0.68%)	0.54
E7	294	3 (1.02%)	98	1 (1.02%)	2 (2.04%)	0.26
E1	1812	12 (0.66%)	604	8 (1.32%)	4 (0.66%)	0.49
E2	1512	6 (0.40%)	384	4 (1.04%)	2 (0.52%)	0.26
E4	534	1 (0.19%)	178	1 (0.56%)	0 (0%)	0.00
L2	1512	16 (1.06%)	504	9 (1.79%)	7 (1.39%)	0.39
L1	1551	8 (0.52%)	517	5 (0.97%)	3 (0.58%)	0.48
LCR	490	1 (0.20%)	–	–	–	–
<b>Total</b>	<b>7186</b>	<b>50 (0.70%)</b>	<b>2431</b>	<b>31 (1.28%)</b>	<b>19 (0.78%)</b>	

HPVX by metagenomics sequencing of cervical DNA [40]. Future studies should examine whether Gamma-6 viruses can be identified in other mucosal and cutaneous sites. In addition to the lack of E6 it is apparent from the phylogenetic tree (Fig. 1) that the four members of the Gamma-6 species, including the novel HPV214 described in this study, have a considerable phylogenetic distance from other gamma HPV types.

The E5 protein is only present in *Alphapapillomaviruses* and absent in Beta, Gamma, Mu and Nu-papillomavirus [41]. The functions of this 90 aa protein in high risk HPVs range from the binding of platelet derived growth factor (PDGF) and hence activating it [42], induction of cell transformation and MHC class 1 antigen presentation inhibition (thereby evading immune response) [43]. However, *Gamma-papillomaviruses* in the absence of E5 proteins are still able to evade the immune system by interfering with IFN $\gamma$  anti-viral pathway [44].

The Gamma-6 species lack the E6 ORF and so did HPV214 [15]. The E6 of high risk HPVs protein is essential in binding of p53 tumour suppressor protein and cell cycle dysregulation [45]. It appears however that this lost function in HPV214 may be compensated for in its E7 protein as it has an LxCxE motif that has been shown to bind pRB in HPV16 and other high risk HPV types, which is not present in the other novel types (see Table 4). It has been shown elsewhere that HPV108, another member of the Gamma-6 species can induce dysplasia in organotypic keratinocytes without having the E6 protein [15], alluding to the fact that once the E6 binding function is not present there are compensatory mechanisms developed through evolution for the viruses to adapt to the host. It was recently shown that gamma-6 HPVs acquired a 37 amino acid protein named E10, the E10 protein is upstream of the E7 start codon [18,19] (Fig. S1a). However the acquisition of E10 does not likely compensate for all E6 functions.

Zinc fingers are finger like zinc binding domains in protein sequences. These are essential in protein to protein interactions and binding to DNA [46]. The E6 and E7 protein of the novel HPVs have

zinc finger binding domains. In high risk HPV types particularly HPV16 the zinc domains facilitate E6 binding to a number of cellular proteins including p53 [45]. In high risk HPV types the binding of p53 by E6 protein constitute, in combination with the E7 protein binding of pRB, the hallmark of HPV carcinogenesis [47]. The E7 protein of most oncogenic HPVs contain a pRB binding domain represented as LxCxE [48], and of the novel types only HPV214 contained this domain.

We found NLS-like sequences in E1, E2, L1 and L2 proteins of all the novel HPV types. In the L1 and L2 proteins they were NLS-like sequences with slight modifications from the KRK and KRRL signatures, as reported in HPV16 [49] and HPV199 [34]. E1 proteins play a primary role in viral DNA replication and hence are found in the nucleus of the host cell. To facilitate this function, E1 proteins have amino acid sequences that are necessary for directing this nuclear localisation [34]. E1 proteins have a bipartite NLS composed of two clusters separated by about 27-30aa (KRK and KRRL), among the novel types they were separated by 28aa. Another conserved NLS motif RKRxR/KRRR/KRKR, previously described in *Alpha*-HPVs [50] and also promotes nuclear localisation, was also found in the E2 proteins of the Gamma-HPVs. NLS motifs were also identified in the C-terminus of the L1 and L2 proteins of the novel Gamma-HPVs with the motif (K/R) $_3$ R(K/R), these have also been described elsewhere [49,51].

Other domains are the PDZ binding domain in the N-terminal of E6 [52], the ATP binding site of E1, the DNA recognition helix of E2 and the furin cleavage site of L2. All these domains play a role in HPV life cycle, the details of the mechanisms of their action are beyond the scope of this paper. However, the transmembrane domain at the N-terminal of the L2 minor capsid protein has been recently described in HPV16 as essential for the translocation of viral DNA across phospholipid bilayer membranes [53,54]. The domain consists of G(x) $_3$ G motifs and similar glycine zippers G(x) $_3$ G(x) $_3$ G motifs that together work in unison to facilitate packing of DNA helices to pass lipid bilayers [34]. This domain was also identified in all the L2 proteins of the novel types,

but its functionality in *Gammapapillomaviruses* has not been explored.

LCR palindromic E2 binding sites ACC-(N)<sub>6</sub>GGT [55] were present in all the novel HPV types, these have also been described elsewhere [56]. The origin of replication of papillomaviruses lies in the LCR and contains more than one E2 binding sites (Table 4) [57]. The E2 protein of HPV16 acts as an activator and repressor of viral transcription and initiation of replication, partitioning of genome and binds to two forms of palindromic sites: ACC-(N)<sub>4</sub>-CGGT and ACC-(N)<sub>6</sub>-GGT. The former site binds with higher affinity compared to the latter [58]. The novel HPVs described in this study had the latter form of the palindromic binding sites which bind weakly as described for HPV16, but this may not apply to the E2 from the novel types.

Polyadenylation sites are adenine rich and facilitate viral mRNA splicing. The early sites are positioned at the 5' end of the L2 protein, while the late polyadenylation sites are usually downstream of the L1 protein within the LCR [17,59]. We identified early polyadenylation sites (-AATAAA-) at the N-terminal (5') end of the L2 protein of all the novel HPV types. The late polyadenylation sites were located in the LCR but at differing positions in each novel type.

In this study we showed that HPV212, HPV213 and HPV215 novel types each had full genome clones with at least 99.5% pairwise identity to each other (Fig. 3). It has been suggested that differences in a single genetic region cannot be used to define a variant, and rather that the complete genome be used for variant classification instead of just the L1 ORF [60]. A common nomenclature for HPV variants and sub-lineages using complete genomes is being implemented [61]. The complete genomes of each of HPV212, HPV213 and HPV215 clones displayed < 0.5% pairwise difference in nucleotide sequences. By strict definition a sub-lineage is an isolate of the same HPV type that differs from the other by a minimal of 0.5% and maximal of 1% difference [61–63] and hence the clones do not fit the definition of sub lineages according to Burk and co-workers [61].

The intra-host genetic diversity of the novel gamma HPVs was examined here by Illumina sequencing of multiple whole genome clones of each type amplified from individual clinical specimens. This is the first study of intra-host variation of *Gammapapillomaviruses*. While there was no variation in the genomes of HPV211, HPV214 and HPV216 we identified 67, 51 and 50 variable nucleotide sites in the genomes of HPV212, HPV213 and HPV215, respectively. This diversity was greater than expected. Papillomaviruses have long been thought to have a low rate of mutation and to co-evolve with their hosts [32,64,65] due to the fact that they hijack the host cellular DNA replication machinery for replication which includes high fidelity polymerases with proof reading activity (with an error rate of about  $4.3 \times 10^{-5}$  [66]) and post-replication repair. However, similar to our findings, several recent NGS-based studies report high intra-host variability in *Alphapapillomaviruses* in clinical specimens used Ion Torrent sequencing of long PCR amplicons of the HPV16 genome and identified between 3 and 175 variable nucleotide sites per genome in samples [65,67,68]. Hirose and co-workers [68] identified an average of 7 nucleotide variations (range 0–85) per sample in the genomes of HPV16, HPV52 and HPV58. Dube Mandishora and co-workers [65] identified hundreds of variant sites in the PGMV region of the L1 gene of HPV16, HPV18 and HPV52 using Illumina sequencing. Several explanations have been proposed for the high variability observed, including the generation of variants during infection due to the activation of host polymerases with lower fidelity [65] as well as the recruitment of members of the APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of mutagenic enzymes [68–70] resulting in mutations rates higher than the human autosomal mutation rate.

The  $d_N/d_S$  ratios for all ORFs of HPV212, HPV213 and HPV215 were less than 1 indicating purifying selection (Table 5). This is in agreement with previous findings that the ORFs of HPV16 and other *Alphapapillomavirus* are under strong purifying selection [71–73]. Purifying selection is likely the result of the requirement for maintaining functional viral proteins for the successful completion of the virus life cycle.

There were no nonsynonymous changes in conserved functional domains except for a Cysteine to Phenylalanine at the last position of the E7 zinc finger binding domain found in two HPV213 clones. Whether this amino acid change is deleterious to the functioning of E7 was not explored.

## 5. Conclusions

The characterisation and classification of HPV211, HPV212, HPV213, HPV214, HPV215 and HPV216 add these novel types to the repertoire of the ever expanding *Gammapapillomavirus* genus. We make the fourth announcement of an HPV lacking the E6 ORF. It is apparent from the phylogenetic tree (Fig. 1) based on L1 nucleotide sequences that the four members of the Gamma-6 species, besides the unusual genome (lack of E6 ORF), also have a considerable phylogenetic distance from other HPV types. Further investigations into the Gamma-6 species and tissue tropism and potential disease association of the other novel HPVs described here is warranted in order to empirically evaluate their clinical significance. We recommend further studies into intra-host viral diversity.

## Declaration of interest statement

The authors declare no commercial or other association that might pose a conflict of interest.

## Acknowledgements

This study was supported in part by the Poliomyelitis Research Foundation (PRF) of South Africa and in part by the South African Research Chairs Initiative of the Department of Science and Technology of South Africa and NRF and in part by the National Research Foundation of South Africa (Grant Number: 64815).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.pvr.2019.02.005>.

## References

- [1] H.U. Bernard, R.D. Burk, Z. Chen, K. van Doorslaer, H. zur Hausen, E.M. de Villiers, Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments, *Virology* 401 (1) (2010) 70–79.
- [2] E.M. de Villiers, C. Fauquet, T.R. Broker, H.U. Bernard, H. zur Hausen, Classification of papillomaviruses, *Virology* 324 (2004) 17–27.
- [3] L.S.A. Mühr, C. Eklund, J. Dillner, Towards quality and order in human papillomavirus research, *Virology* 519 (2018) 74–76.
- [4] A.E. Ure, O. Forslund, Characterization of human papillomavirus type 154 and tissue tropism of gammapapillomaviruses, *PLoS One* 9 (2014) e89342.
- [5] Y. Ma, R. Madupu, U. Karaoz, C.W. Nossa, L. Yang, S. Yooseph, P.S. Yachimski, E.L. Brodie, K.E. Nelson, Z. Pei, Human papillomavirus community in healthy persons, defined by metagenomics analysis of human microbiome project shotgun sequencing data sets, *J. Virol.* 88 (2014) 4786–4797.
- [6] A. Antonsson, O. Forslund, H. Ekberg, G. Sterner, B.G. Hansson, The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensal nature of these viruses, *J. Virol.* 74 (2000) 11636–11641.
- [7] J. Ekström, D. Bzhalava, D. Svenback, O. Forslund, J. Dillner, High throughput sequencing reveals diversity of Human Papillomaviruses in cutaneous lesions, *Int. J. Cancer* 129 (2011) 2643–2650.
- [8] L. Sichero, C.M. Pierce Campbell, S. Ferreira, J.S. Sobrinho, M. Luiza Baggio, L. Galan, R.C. Silva, E. Lazcano-Ponce, A.R. Giuliano, L.L. Villa, Broad HPV distribution in the genital region of men from the HPV infection in men (HIM) study, *Virology* 443 (2013) 214–217.
- [9] L. Sichero, C.M. Pierce Campbell, W. Fulp, S. Ferreira, J.S. Sobrinho, M. Baggio, L. Galan, R.C. Silva, E. Lazcano-Ponce, A.R. Giuliano, L.L. Villa, High genital prevalence of cutaneous human papillomavirus DNA on male genital skin: the HPV Infection in Men Study, *BMC Infect. Dis.* 14 (2014) 677.
- [10] D. Bottalico, Z. Chen, A. Dunne, J. Ostolozza, S. McKinney, C. Sun, N.F. Schlecht, M. Fatahzaheh, R. Herrero, M. Schiffman, R.D. Burk, The oral cavity contains abundant known and novel human papillomaviruses from the Betapapillomavirus and Gammapapillomavirus genera, *J. Infect. Dis.* 204 (2011) 787–792.
- [11] O. Forslund, H. Johansson, K.G. Madsen, K. Kofoed, The nasal mucosa contains a



- large spectrum of human papillomavirus types from the Betapapillomavirus and Gammapapillomavirus genera, *J. Infect. Dis.* 208 (2013) 1335–1341.
- [12] L. Slichero, A.G. Nyitray, E.M. Nunes, B. Nepal, S. Ferreira, J.S. Sobrinho, M.L. Baggio, L. Galan, R.C. Silva, E. Lazcano-Ponce, A.R. Giuliano, L.L. Villa, Diversity of human papillomavirus in the anal canal of men: the HIM Study, *Clin. Microbiol. Infect.* 21 (2015) 502–509.
- [13] D. Bzhalava, P. Guan, S. Franceschi, J. Dillner, G. Clifford, A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types, *Virology* 445 (2013) 224–231.
- [14] M. Grace, K. Munger, Proteomic analysis of the gamma human papillomavirus type 197 E6 and E7 associated cellular proteins, *Virology* 500 (2017) 71–81.
- [15] R.J. Nobre, E. Herrera-Hernandez, J.W. Fei, L. Langbein, S. Kaden, H.J. Grone, E.M. de Villiers, E7 oncoprotein of novel human papillomavirus type 108 lacking the E6 gene induces dysplasia in organotypic keratinocyte cultures, *J. Virol.* 83 (2009) 2907–2916.
- [16] A. Ameur, T.L. Meiring, I. Bunikis, S. Haggqvist, C. Lindau, J.H. Lindberg, I. Gustavsson, Z.Z. Mbulawa, A.L. Williamson, U. Gyllensten, Comprehensive profiling of the vaginal microbiome in HIV positive women using massive parallel semiconductor sequencing, *Sci. Rep.* 4 (2014) 4398.
- [17] Z. Chen, M. Schiffman, R. Herrero, R. Desalle, R.D. Burk, Human papillomavirus (HPV) types 101 and 103 isolated from cervicovaginal cells lack an E6 open reading frame (ORF) and are related to gamma-papillomaviruses, *Viol. J.* 360 (2007) 447–453.
- [18] K. Van Doorslaer, A.A. McBride, Molecular archeological evidence in support of the repeated loss of a papillomavirus gene, *Sci. Rep.* 6 (2016) 33028.
- [19] K. Van Doorslaer, Z. Li, S. Xirasagar, P. Maes, D. Kaminsky, D. Liou, Q. Sun, R. Kaur, Y. Huyen, A.A. McBride, The Papillomavirus Episteme: a major update to the papillomavirus sequence database, *Nucleic Acids Res.* 45 (2017) 499–506.
- [20] T.L. Meiring, Z.Z.A. Mbulawa, M. Lesosky, D. Coetzee, A.L. Williamson, High diversity of alpha, beta and gamma human papillomaviruses in genital samples from HIV-negative and HIV-positive heterosexual South African men, *Papillomavirus Res.* 3 (2017) 160–167.
- [21] O. Forslund, A. Antonsson, P. Nordin, B. Stenquist, B.G. Hansson, A broad range of human papillomavirus types detected with a general PCR method suitable for analysis of cutaneous tumours and normal skin, *J. Gen. Virol.* 80 (Pt 9) (1999) 2437–2443.
- [22] Z.Z. Mbulawa, D. Coetzee, D.J. Marais, M. Kamupira, E. Zwane, B. Allan, D. Constant, J.R. Moodley, M. Hoffman, A.L. Williamson, Genital human papillomavirus prevalence and human papillomavirus concordance in heterosexual couples are positively associated with human immunodeficiency virus coinfection, *J. Infect. Dis.* 199 (10) (2009) 1514–1524.
- [23] M.G. Reese, F.H. Eeckman, D. Kulp, D. Haussler, Improved splice site detection in Genie, *J. Comput. Biol.* 4 (1997) 311–323.
- [24] K. Van Doorslaer, Q. Tan, S. Xirasagar, S. Bandaru, V. Gopalan, Y. Mohamoud, Y. Huyen, A.A. McBride, The Papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis, *Nucleic Acids Res.* 41 (2013) D571–D578.
- [25] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinf.* 5 (2004) 113.
- [26] S. Guindon, J.F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel, New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0, *Syst. Biol.* 59 (2010) 307–321.
- [27] D. Darriba, G.L. Taboada, R. Doallo, D. Posada, jModelTest 2: more models, new heuristics and parallel computing, *Nat. Methods* 9 (2012) 772.
- [28] M. Anisimova, O. Gascuel, Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative, *Syst. Biol.* 55 (2006) 539–552.
- [29] I. Letunic, P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees, *Nucleic Acids Res.* 44 (2016) W242–W245.
- [30] K.B. Allen G. Rodrigo, Gerald H. Learn (Eds.), *Alternatives: HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences*, vol. 4, Kluwer Academic Publishers, Dordrecht, Netherlands, 2000, pp. 55–72.
- [31] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, *Mol. Biol. Evol.* 3 (1986) 418–426.
- [32] Z. Chen, R. DeSalle, M. Schiffman, R. Herrero, R.D. Burk, Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97, *J. Virol.* 83 (2009) 1443–1455.
- [33] J. Li, H. Cai, Z. Xu, Q. Wang, D. Hang, N. Shen, M. Liu, C. Zhang, A. Abliz, Y. Ke, Nine complete genome sequences of cutaneous human papillomavirus genotypes isolated from healthy skin of individuals living in rural He Nan province, China, *J. Virol.* 86 (2012) 11936.
- [34] A. Ostrbenk, B.J. Kocjan, L. Hosnjak, J. Li, Q. Deng, A. Sterbenc, M. Poljak, Identification of a novel human papillomavirus, type HPV199, isolated from a nasopharynx and anal canal, and complete genomic characterization of papillomavirus species gamma-12, *PLoS One* 10 (2015) e0138628.
- [35] T. Mitsuishi, I. Ohsawa, T. Kato, N. Egawa, T. Kiyono, Molecular cloning and characterisation of a novel type of human papillomavirus 160 isolated from a flat wart of an immunocompetent patient, *PLoS One* 8 (2013) e79592.
- [36] D. Botalicco, Z. Chen, A. Dunne, J. Ostolza, S. McKinney, C. Sun, N.F. Schlecht, M. Fatahzadeh, R. Herrero, M. Schiffman, R.D. Burk, The oral cavity contains abundant known and novel human papillomaviruses from the Betapapillomavirus and Gammapapillomavirus genera, *J. Infect. Dis.* 204 (2011) 787–792.
- [37] E. Sturegard, H. Johansson, J. Ekstrom, B.G. Hansson, A. Johnsson, E. Gustafsson, J. Dillner, O. Forslund, Human papillomavirus typing in reporting of condyloma, *Sex. Transm. Dis.* 40 (2013) 123–129.
- [38] A. Kohler, M. Gottschling, K. Manning, M.D. Lehmann, E. Schulz, D. Kruger-Corcoran, E. Stockfleth, I. Nindl, Genomic characterization of ten novel cutaneous human papillomaviruses from keratotic lesions of immunosuppressed patients, *J. Gen. Virol.* 92 (2011) 1585–1594.
- [39] M.E. McLaughlin-Drubin, K. Munger, Oncogenic activities of human papillomaviruses, *Virus Res.* 143 (2) (2009) 195–208.
- [40] A. Ameur, T.L. Meiring, I. Bunikis, S. Haggqvist, C. Lindau, J.H. Lindberg, I. Gustavsson, Z.Z. Mbulawa, A.L. Williamson, U. Gyllensten, Comprehensive profiling of the vaginal microbiome in HIV positive women using massive parallel semiconductor sequencing, *Sci. Rep.* 4 (2014) 4398.
- [41] A. Venuti, F. Paolini, L. Nasir, A. Corteggio, S. Roperto, M.S. Campo, G. Borzacchiello, Papillomavirus E5: the smallest oncoprotein with many functions, *Mol. Canc.* 10 (2011) 140.
- [42] L. Petti, L.A. Nilson, D. DiMaio, Activation of the platelet-derived growth factor receptor by the bovine papillomavirus E5 transforming protein, *EMBO J.* 10 (1991) 845–855.
- [43] M.A. Stanley, M.R. Pett, N. Coleman, HPV: from infection to cancer, *Biochem. Soc. Trans.* 35 (2007) 1456–1460.
- [44] P. Cordano, V. Gillan, S. Bratlie, V. Bouvard, L. Banks, M. Tommasino, M.S. Campo, The E6E7 oncoproteins of cutaneous human papillomavirus type 38 interfere with the interferon pathway, *Viol. J.* 377 (2008) 408–418.
- [45] N.A. Wallace, D.A. Galloway, Novel functions of the human papillomavirus E6 oncoproteins, *Annu. Rev. Virol.* 2 (2015) 403–423.
- [46] M. Wayengera, Zinc finger arrays binding human papillomavirus types 16 and 18 genomic DNA: precursors of gene-therapeutics for in-situ reversal of associated cervical neoplasia, *Theor. Biol. Med. Model.* 9 (2012) 30.
- [47] K. Munger, A. Baldwin, K.M. Edwards, H. Hayakawa, C.L. Nguyen, M. Owens, M. Grace, K. Huh, Mechanisms of human papillomavirus-induced oncogenesis, *J. Virol.* 78 (2004) 11451–11460.
- [48] J. Doorbar, N. Egawa, H. Griffin, C. Kranjec, I. Murakami, Human papillomavirus molecular biology and disease association, *Rev. Med. Virol.* 25 (Suppl 1) (2015) 2–23.
- [49] J. Zhou, J. Doorbar, X.Y. Sun, L.V. Crawford, C.S. McLean, I.H. Frazer, Identification of the nuclear localization signal of human papillomavirus type 16 L1 protein, *Virology* 185 (1991) 625–632.
- [50] N. Zou, B.Y. Lin, F. Duan, K.Y. Lee, G. Jin, R. Guan, G. Yao, E.J. Lefkowitz, T.R. Broker, L.T. Chow, The hinge of the human papillomavirus type 11 E2 protein contains major determinants for nuclear localization and nuclear matrix association, *J. Virol.* 74 (2000) 3761–3770.
- [51] L.M. Nelson, R.C. Rose, L. LeRoux, C. Lane, K. Bruya, J. Moroiianu, Nuclear import and DNA binding of human papillomavirus type 45 L1 capsid protein, *J. Cell. Biochem.* 79 (2000) 225–238.
- [52] E.M. Bolatti, D. Chouhy, P.E. Casal, G.R. Perez, E.J. Stella, A. Sanchez, M. Gorosito, R.F. Bussy, A.A. Giri, Characterization of novel human papillomavirus types 157, 158 and 205 from healthy skin and recombination analysis in genus gamma-Papillomavirus, *Infect. Genet. Evol.* 42 (2016) 20–29.
- [53] M.P. Bronnimann, J.A. Chapman, C.K. Park, S.K. Campos, A transmembrane domain and GxxxG motifs within L2 are essential for papillomavirus infection, *J. Virol.* 87 (2013) 464–473.
- [54] J.W. Wang, R.B. Roden, L2, the minor capsid protein of papillomavirus, *Virology* 445 (2013) 175–186.
- [55] L. Li, P. Barry, E. Yeh, C. Glaser, D. Schnurr, E. Delwart, Identification of a novel human gammapapillomavirus species, *J. Gen. Virol.* 90 (2009) 2413–2417.
- [56] C.D. Newhouse, S.J. Silverstein, Orientation of a novel DNA binding site affects human papillomavirus-mediated transcription and replication, *J. Virol.* 75 (2001) 1722–1735.
- [57] F. Sverdrup, S.A. Khan, Two E2 binding sites alone are sufficient to function as the minimal origin of replication of human papillomavirus type 18 DNA, *J. Virol.* 69 (1995) 1319–1323.
- [58] C.L. Bedrosian, D. Bastia, The DNA-binding domain of HPV-16 E2 protein interaction with the viral enhancer: protein-induced DNA bending and role of the non-conserved core sequence in binding site affinity, *Virology* 174 (1990) 557–575.
- [59] Z. Chen, M. Schiffman, R. Herrero, R.D. Burk, Identification and characterization of two novel human papillomaviruses (HPVs) by overlapping PCR: HPV102 and HPV106, *J. Gen. Virol.* 88 (2007) 2952–2955.
- [60] Z. Chen, L.B. de Freitas, R.D. Burk, Evolution and classification of oncogenic human papillomavirus types and variants associated with cervical cancer, *Methods Mol. Biol.* 1249 (2015) 3–26.
- [61] R.D. Burk, A. Harari, Z. Chen, Human papillomavirus genome variants, *Virology* 445 (2013) 232–243.
- [62] I.E. Calleja-Macias, M. Kalantari, B. Allan, A.L. Williamson, L.P. Chung, R.J. Collins, R.E. Zuna, S.T. Dunn, R. Ortiz-Lopez, H.A. Barrera-Saldana, H.A. Cubie, K. Cuschieri, L.L. Villa, H.U. Bernard, Papillomavirus subtypes are natural and old taxa: phylogeny of human papillomavirus types 44 and 55 and 68a and -b, *J. Virol.* 79 (2005) 6565–6569.
- [63] H.U. Bernard, I.E. Calleja-Macias, S.T. Dunn, Genome variation of human papillomavirus types: phylogenetic and medical implications, *Int. J. Cancer* 118 (2006) 1071–1076.
- [64] I.G. Bravo, M. Felez-Sanchez, Papillomaviruses: viral evolution, cancer and evolutionary medicine, *Evol. Med. Publ. Health* (2015) 32–51 (2015).
- [65] R.S. Dube Mandishora, K.S. Gjotterud, S. Lagstrom, B. Stray-Pedersen, K. Duri, N. Chin'ombe, M. Nygard, I.K. Christiansen, O.H. Ambrur, M.Z. Chirenje, T.B. Rounge, Intra-host sequence variability in human papillomavirus, *Papillomavirus Res.* 5 (2018) 180–191.
- [66] D.A. Korona, K.G. Lecompte, Z.F. Pursell, The high fidelity and unique error signature of human DNA polymerase epsilon, *Nucleic Acids Res.* 39 (2011) 1763–1773.
- [67] C.M. de Oliveira, I.G. Bravo, N.C. Santiago e Souza, M.L. Genta, J.H. Fregiani,

- M. Tacla, J.P. Carvalho, A. Longatto-Filho, J.E. Levi, High-level of viral genomic diversity in cervical cancers: a Brazilian study on human papillomavirus type 16, *Infect. Genet. Evol.* 34 (2015) 44–51.
- [68] Y. Hirose, M. Onuki, Y. Tenjimbayashi, S. Mori, Y. Ishii, T. Takeuchi, N. Tasaka, T. Satoh, T. Morisada, T. Iwata, S. Miyamoto, K. Matsumoto, A. Sekizawa, I. Kukimoto, Within-host variations of human papillomavirus reveal APOBEC signature mutagenesis in the viral genome, *J. Virol.* (2018) 92.
- [69] L. Mirabello, M. Yeager, K. Yu, G.M. Clifford, Y. Xiao, B. Zhu, M. Cullen, J.F. Boland, N. Wentzensen, C.W. Nelson, T. Raine-Bennett, Z. Chen, S. Bass, L. Song, Q. Yang, M. Steinberg, L. Burdett, M. Dean, D. Roberson, J. Mitchell, T. Lorey, S. Franceschi, P.E. Castle, J. Walker, R. Zuna, A.R. Kreimer, D.C. Beachler, A. Hildesheim, P. Gonzalez, C. Porras, R.D. Burk, M. Schiffman, HPV16 E7 genetic conservation is critical to carcinogenesis, *Cell* 170 (2017) 1164–1174 e1166.
- [70] J.P. Vartanian, D. Guetard, M. Henry, S. Wain-Hobson, Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions, *Science* 320 (2008) 230–233.
- [71] V.R. DeFilippis, F.J. Ayala, L.P. Villarreal, Evidence of diversifying selection in human papillomavirus type 16 E6 but not E7 oncogenes, *J. Mol. Evol.* 55 (2002) 491–499.
- [72] A. Carvajal-Rodriguez, Detecting recombination and diversifying selection in human alpha-papillomavirus, *Infect. Genet. Evol.* 8 (2008) 689–692.
- [73] Z. Chen, M. Terai, L. Fu, R. Herrero, R. DeSalle, R.D. Burk, Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses, *J. Virol.* 79 (2005) 7014–7023.
- [74] D. Bzhalava, C. Eklund, J. Dillner, International standardization and classification of human papillomavirus types, *Virology* (2015) 2341–2344 2015 Jan 2018;2476C.