HHS Public Access

Author manuscript

Virology. Author manuscript; available in PMC 2022 July 14.

Published in final edited form as:

Virology. 2018 July ; 520: 1–10. doi:10.1016/j.virol.2018.04.017.

Generation of a novel next-generation sequencing-based method for the isolation of new human papillomavirus types

Rosario N. Brancaccio[a,1], **Alexis Robitaille**[a,1], **Sankhadeep Dutta**[a], **Cyrille Cuenin**[a], **Daiga Santare**[b], **Girts Skenders**[b], **Marcis Leja**[b], **Nicole Fischer**[c,d], **Anna R. Giuliano**[e], **Dana E. Rollison**[e,f], **Adam Grundhoff**[d,g], **Massimo Tommasino**[a,*], **Tarik Gheit**[a,*]

[a]International Agency for Research on Cancer, Lyon, France

[b]Institute of Clinical and Preventive Medicine, University of Latvia, Riga, Latvia

[c]Institute for Medical Microbiology, Virology and Hygiene, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[d]German Center for Infection Research, partner site Hamburg Borstel Lubeck, Riems, Germany

[e]Center for Infection Research in Cancer, Moffitt Cancer Center, Tampa, FL, USA

[f]Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL, USA

[g]Heinrich-Pette Institute, Leibniz Institute for Experimental Virology, Hamburg Germany

Abstract

With the advent of new molecular tools, the discovery of new papillomaviruses (PVs) has accelerated during the past decade, enabling the expansion of knowledge about the viral populations that inhabit the human body. Human PVs (HPVs) are etiologically linked to benign or malignant lesions of the skin and mucosa. The detection of HPV types can vary widely, depending mainly on the methodology and the quality of the biological sample. Next-generation sequencing is one of the most powerful tools, enabling the discovery of novel viruses in a wide range of biological material. Here, we report a novel protocol for the detection of known and unknown HPV types in human skin and oral gargle samples using improved PCR protocols combined with next-generation sequencing. We identified 105 putative new PV types in addition to 296 known types, thus providing important information about the viral distribution in the oral cavity and skin.

Keywords

Broad-spectrum HPV PCR primers; Next-generation sequencing; New papillomaviruses

[*]Corresponding authors. icb@iarc.fr (M. Tommasino), gheitt@iarc.fr (T. Gheit).

[1]These authors contributed equally to this work.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2018.04.017.

1. Introduction

Human papillomaviruses (HPVs) are non-enveloped viruses with double-stranded circular DNA of about 8 kb that can colonize the mucosal and cutaneous epithelia (Bernard et al., 2010; Bzhalava et al., 2013). To date, more than 200 PVs have been isolated from different body sites and fully characterized, and this number continues to grow (Bzhalava et al., 2015; Smelov et al., 2017). Based on the nucleotide sequences of the major capsid protein L1, HPVs are classified into genera, species, and types (Bernard et al., 2010). HPV types are organized into five major genera: alpha, beta, gamma, mu, and nu (de Villiers et al., 2004). The genera alpha, beta, and gamma include the majority of the known HPVs. The alpha HPV types have been extensively studied, because of their clear association with human carcinogenesis (Tommasino, 2014). The high-risk (HR) HPV group includes at least 12 HPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59), which are the etiological agents of anogenital cancers and a subset of head and neck cancers, particularly oropharyngeal cancer (Bouvard et al., 2009; Haedicke and Iftner, 2013). The genus alpha also includes the low-risk HPV types (HPV6 and 11) that are associated with benign genital lesions and with laryngeal disease in children (Giuliano et al., 2008b; Goon et al., 2008).

The genus beta includes approximately 50 different HPV types, fully characterized, that are subdivided into five species (beta HPV species 1–5). The majority of the beta HPV types belong to species beta-1 and beta-2 and are widely present in the skin of healthy individuals. Only 7 HPV types have been classified into the species beta-3 ($n = 4$), beta-4 ($n = 1$), and beta-5 ($n = 2$). HPV types of genus beta can induce warts and have been associated with certain forms of non-melanoma skin carcinoma (NMSC) (Orth, 2006). The first beta HPVs, HPV5 and 8, were isolated from the skin of patients with epidermodysplasia verruciformis (EV), a rare autosomal recessive hereditary skin disorder that confers high susceptibility to beta HPV infection and cutaneous squamous cell carcinoma development at sun-exposed regions (Pfister, 2003). Several studies showed that beta HPV types are associated with NMSC development in non-EV individuals (Andersson et al., 2008; Berkhout et al., 2000; Bouwes Bavinck et al., 2010; Casabonne et al., 2007; Cornet et al., 2012; de Jong-Tieben et al., 1995; Harwood et al., 2000; Iannacone et al., 2014; Iftner et al., 2003; Karagas et al., 2006; Waterboer et al., 2008). Patients with a history of NMSC show elevated positivity for markers of beta HPV infection compared with healthy individuals (Ally et al., 2013; Asgari et al., 2008; Iannacone et al., 2012). Recent studies reported the presence of beta HPV types at additional anatomical sites other than the skin, such as the oral mucosal epithelium, eyebrow hairs, penile and external genital samples, and the anal canal (Arroyo et al., 2013; Barzon et al., 2011; Donà et al., 2016; Pierce Campbell et al., 2016; Smelov et al., 2017).

Species beta-3 HPV types appear to have a dual tropism, being present in the skin and the mucosal epithelia (Forslund et al., 2013; Hampras et al., 2017). Interestingly, studies in *in vitro* and *in vivo* experimental models have highlighted some biological similarities between beta-3 HPV and mucosal HR HPV types (Cornet et al., 2012; Viarisio et al., 2016). In addition, Viarisio et al. (2016) showed that beta-3-HPV49 transgenic mice were highly susceptible to upper digestive tract carcinogenesis upon initiation with 4-nitroquinoline 1-oxide.

HPVs from the gamma, mu, and nu genera induce cutaneous papillomas or warts (de Villiers et al., 2004) and have been poorly investigated so far. To date, approximately 80 different gamma HPV types have been isolated from the skin and genital tract (retrieved from GenBank, September 2017).

In addition to the fully characterized HPV types, a substantial number of partial genomic sequences of putative novel HPV types have been deposited to GenBank, indicating that many more HPV types exist. So far, the molecular biology techniques for the isolation of novel HPV types have been based mainly on the use of degenerate and/or consensus primers, followed by cloning and Sanger sequencing (Chouhy et al., 2010; Forslund et al., 1999). However, considering the large number of recently characterized HPV genomes, degenerate primers may be improved in order to discover novel HPV types. In particular, this strategy may lead to the expansion of species that so far include a very small number of HPV types, such as species beta-3 ($n = 4$), beta-4 ($n = 1$), and beta-5 ($n = 2$).

In this study, we used novel and well-validated consensus and degenerate primers to amplify genomic HPV sequences from human DNA isolated from oral and skin specimens. Analysis of the PCR products by next-generation sequencing (NGS) resulted in the identification of 105 putative new PV types.

2. Materials and methods

2.1. Sample collection and DNA extraction

Skin swabs and oral rinses from two different ongoing studies aiming to determine the prevalence of viral DNA and its associations with disease were used in the present analysis (Hampras et al., 2014, 2015; Nunes et al., 2016; Pierce Campbell et al., 2013, 2016).

Skin swab specimens ($n = 119$) were randomly selected from the VIRUSCAN Study, an ongoing five-year (2014–2019) prospective cohort study conducted at Moffitt Cancer Center and the University of South Florida (R01CA177586-01; “Prospective study of cutaneous viral infections and non-melanoma skin cancer”). An area of approximately 5×5 cm of the top of the sun-exposed forearm was sprayed with 0.9% saline solution. A cotton-tipped Dacron swab (Digene, Gaithersburg, MD, USA) was then rubbed back and forth a few times to collect exfoliated skin cells. Individual swabs were placed in a separate vial and preserved in Digene Standard Transport Medium.

In addition, 62 oral rinses were randomly selected from the HPV Infection in Men (HIM) study, a large, multinational (Brazil, Mexico, and the USA) prospective cohort study of the natural history of HPV infection in men. The HIM study methods have previously been described in detail (Giuliano et al., 2008a, 2009, 2011; Nyitray et al., 2011). A further 85 oral samples were selected from a pilot study that aimed to estimate the prevalence of *Helicobacter pylori* in oral gargles from a Latvian population. The study was approved (No. 8-A/15) by the Ethics Committee of Riga East University Hospital Support Foundation.

After DNA extraction, all samples were analyzed at the International Agency for Research on Cancer (Lyon, France) for viral DNA from HPV.

2.2. PCR protocols

The following PCR protocols using different sets of primers were run (Table 1): (i) CUT primers, as previously described (Chouhy et al., 2010); (ii) FA-type (FAP) primers, as previously described (Forslund et al., 1999); (iii) a new set of FAP primers, i.e. FAP59.1, FAP59.2, and FAP64.1 (Fig. 1; Table 1); these primers were used to generate two different primer mixtures (FAPM1 and FAPM2); the PCR conditions were the same as for the original FAP protocol; (iv) a set of 11 beta-3 specific primers (henceforth referred to as beta-3-1) (Table 1); and (v) a set of 4 broad-spectrum beta-3 degenerate primers (henceforth referred to as beta-3-2). The beta-3-1 and beta-3-2 primers were synthesized by MWG Biotech (Ebersberg, Germany) and mixed to obtain a 10 × solution containing 2 μM of each primer. PCR was performed with the Qiagen Multiplex PCR kit (Hilden, Germany) according to the manufacturer's instructions. The use of these primers enables the amplification of a region in the L1 gene of approximately 450 bp.

2.3. Validation of the new set of primers

To evaluate the sensitivity of the novel HPV PCR protocols (beta-3-1, beta-3-2, FAPM1, and FAPM2), we used an artificial mixture containing cloned HPV genomes at different relative concentrations (10-fold dilution series starting from 10,000 to 0 copies of the viral genome) and mixed with human genomic DNA. PCR products were analyzed by electrophoresis on a 2% agarose gel.

2.4. NGS analysis

The PCR products were purified on a 2% agarose gel using the QIAquick gel extraction kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. One additional purification step was performed to remove any remaining contaminants using the Agencourt AMPure XP PCR purification kit with a beads ratio of 1.8 × (Beckman Coulter). Purified PCR products were divided into eight different pools. Each pool included approximately 50 different amplicons generated from different PCR protocols (Table 2).

Libraries were prepared using the Nextera XT DNA Library preparation kit (Illumina, San Diego, CA, USA). Illumina MiSeq dual-indexed adapters (Illumina, San Diego, CA, USA) were added to each of the PCR pools.

NGS was performed using the Illumina MiSeq kit v3 (600 cycles) on the Illumina MiSeq system. In order to enrich the diversity of the libraries, 10% of PhiX (Illumina, San Diego, California, USA) was added to the NGS reaction.

2.5. Bioinformatics analysis

Quality control was conducted using FastQC (Andrews, 2010) (v0.11.5) and MultiQC (Käller and Ewels, 2016) (v1.0). Trim Galore (v0.4.4) (Krueger, 2015) was used to remove remaining adapter sequences and trim low-quality ends of reads. The merging of forward and reverse reads, the de-replication step, the de novo chimeric sequence identification, and the clustering steps were carried out using VSEARCH (Mahé and Rognes, 2016) (v2.4.0). MegaBlast in the Blast package (v2.6.0+) (Altschul et al., 1990) was launched

against the nucleotide collection (nr/nt, March 2017) database in a local server to enable the identification of the previously constructed clusters.

Another level of clustering was applied for the reads having the same best MegaBlast results inside each pool (based on the E-value). Each cluster of reads was then processed using the CAP3 program in order to assemble contigs (Huang and Madan, 1999).

A reference species phylogenetic tree was constructed based on the full-L1 ORF nucleotide sequences of 458 available PV genomes retrieved from the PaVE database (https://pave.niaid.nih.gov/) (Van Doorslaer et al., 2013) in January 2018. The sequences were aligned at the nucleotide level using the MUSCLE algorithm, with the default parameters (Edgar, 2004), in MEGA7 (Kumar et al., 2016). The final full-length L1-ORF alignment encompassed 458 full L1-ORF nucleotide sequences, 2259 positions, and 627 distinct alignment patterns. MEGA7 was used to test the best substitution model and for the phylogenetic inference. The codon positions included were 1st + 2nd + 3rd + non-coding. Based on the alignment using MUSCLE, all positions with < 95% site coverage were eliminated (partial deletions), to enable the inclusion of taxa with some missing data. There was a total of 1383 positions in the final dataset.

A discrete gamma distribution was used to model evolutionary rate differences among sites (five categories; +G, parameter = 1.0326). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 2.5307% sites).

The initial trees for the heuristic search were obtained automatically by applying the neighbor-joining (NJ)/BioNJ algorithm to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach and then by selecting the topology with the highest log likelihood value (−389774.5274).

Phylogenetic inference was performed with MEGA7 using the general time reversible (GTR) model of nucleotide substitution and 500 bootstrap replicates (Nei and Kumar, 2000).

PaPaRa (v2.5) (Berger and Stamatakis, 2011) was used to align the sequences reconstructed using the CAP3 algorithm with respect to the reference multiple sequence alignment. Subsequently, the evolutionary placement algorithm (EPA) in RAxML (v8.2.11) (Berger et al., 2011; Stamatakis, 2014) was run to place the sequences into the reference species phylogenetic tree. The EPA was run using the same nucleotide substitution model used to infer the reference phylogenetic tree. A script was developed in-house to parse the output format (Matsen et al., 2012) of the EPA.

In addition, a blastn local alignment query of the contigs was used to align them against a comprehensive database of reference PVs present in the PaVE database (n = 330 genomes). This approach mimics locally the L1 taxonomic tool of the PaVE database.

All the results in this study are based on the identification of the sequences using the EPA in RAxML (henceforth referred to as RAxML-EPA). Only the longest sequence was considered for RAxML-EPA classification when several singlets or contigs were available. Krona (Ondov et al., 2011) was used for the graphical representation of the data.

3. Results

3.1. Design and validation of novel HPV PCR primers

As a first step, we generated new consensus primers considering all known beta HPV types. HPV beta-3 species primers were designed by aligning the L1 open reading frame (ORF) from all four beta-3 HPV types (HPV49, 75, 76, and 115) using the ClustalW2 multiple sequence alignment tool (Chenna et al., 2003). Two different sets of primers were generated: (i) a set of 11 specific primers, termed beta-3-1, and (ii) a set of 4 degenerate primers, termed beta-3-2. The composition of the two primer mixtures is shown in Table 1.

As a second approach, we generated additional degenerate primers based on the well-validated FAP primers (Forslund et al., 1999). The FAP primers were developed in 1999 by aligning 77 L1 ORF sequences from different genera that included at that time only a limited number of beta (n = 22) and gamma (n = 5) HPV type sequences, obtained from the 1996 and 1997 HPV Sequence Database Compendia (Myers et al., 1996, 1997).

Forty-six L1 sequences representative of the beta HPV types known to date (Van Doorslaer et al., 2017) were aligned against FAP primer sequences, using the MUSCLE (3.8) multiple sequence alignment tool (Edgar, 2004). Subsequently, three improved broad-spectrum FAP primers, with an increased specificity for beta HPV types, were generated (Fig. 1): FAP59.1, FAP59.2, and FAP64.1. These were mixed in different combinations, generating two different mixtures: FAPM1 and FAPM2 (Table 1). The beta-3 protocol enabled the detection of beta-3 HPVs with a limit of detection of 10 copies. The detection limit using the FAPM1 mixture was 10 copies for HPV types that belong to species beta-2, beta-3, and beta-4 and 1000 copies for beta-5. Using the FAPM2 protocol, the detection limit was 10 copies for HPV types that belong to beta-3, beta-4, and beta-5; however, a lower sensitivity was observed for species beta-2 (10,000 copies) (data not shown).

3.2. NGS data analysis: characterization and taxonomic classification

Randomly selected DNA extracted from skin swabs (n = 119) and oral gargles (n = 147) obtained from healthy individuals was amplified using the PCR protocols described above and the original FAP and CUT protocols (Table 2) (Chouhy et al., 2010; Forslund et al., 1999). PCR products were mixed to obtain 8 different pools (Table 2) and sequenced using the Illumina MiSeq sequencing platform.

A total of 50,017,076 paired-end raw reads were obtained from the NGS analysis. After quality trimming, de-replication, and chimeric PCR sequence removal, 47.3% (23,647,656) of the reads were considered for further analysis. Approximately 67% (16,043,298 reads) were related to PV sequences. Each read was matched against National Center for Biotechnology Information (NCBI) database sequences (nr/nt, March 2017) using the MegaBlast algorithm and assigned to its closest PV type, before contig construction.

Analysis of the data with RAxML-EPA revealed that the reads generated from the sequencing of the 119 skin DNA samples were assigned to a total of 265 different PV types (Fig. 2A; Table S1), which belong mainly to the alpha (33.4%) and beta (29.5%) genera, thus representing the PV distribution in skin. In addition, a substantial fraction of

reads (12.9%) was assigned to taxonomically unclassified PV sequences (hereafter called "unclassified PVs"): bovine papillomavirus type 19 (BPV19), equine papillomavirus type 8 (EcPV8), *Myotis ricketti* papillomavirus 1 (MrPV1), *Pudu puda* papillomavirus type 1 (PpuPV1), and *Sparus aurata* papillomavirus type 1 (SaPV1). Moreover, 9% of the reads were assigned to the gamma genus.

The FAPM1 protocol enabled the detection of 107 PVs (8 alpha, 37 beta, 60 gamma, and 2 mu), and the CUT and FAP protocols enabled the detection of 118 PVs (11 alpha, 36 beta, 68 gamma, 2 mu, and 1 nu) and 87 PVs (3 alpha, 34 beta, 49 gamma, and 1 mu), respectively. The combined beta-3-1 and beta-3-2 protocols generated a majority of reads assigned to a non-human alpha PV type: *Colobus guereza* monkey papillomavirus type 1 (CgPV1). Two reads were assigned to HPV16. Five beta HPVs were detected using these combined protocols (797,800 reads), of which 3 were assigned to species beta-3. Only 2 non-referenced gamma HPV types were detected (HPV-mDysk1 – KX781280 and HPV-mDysk6 – KX781285).

The reads generated from the sequencing of the 147 oral DNA samples were assigned to a total of 161 different PV types. PV types that belong to the genus beta were most common (29.5%), followed by genus gamma (19.6%) and genus alpha (7.8%) (Fig. 2B; Table S2). In addition, a substantial fraction of reads (36.9%) was assigned to taxonomically unclassified PVs: EcPV8, *Miniopterus schreibersii* papillomavirus type 1 (MscPV1), PpuPV1, and SaPV1 (Fig. 2B; Table S2).

The FAPM1 and FAPM2 protocols enabled the detection of 55 PVs (4 alpha, 30 beta, and 21 gamma) and 42 PVs (5 alpha, 21 beta, and 16 gamma), respectively. Forty-six PVs (6 alpha, 17 beta, and 23 gamma) were detected using the CUT protocol (Fig. 2B; Table S2).

Substantial numbers of reads identified in both skin (745,860 reads) and oral (163,448 reads) samples were related to taxonomically classified non-human PVs (i.e. PVs not belonging to the genera alpha, beta, gamma, mu, and nu) (Tables S1 and S2; Fig. 2).

3.3. Subdivision of the NGS reads into known and putative novel PVs

The NGS sequences were divided into two groups, on the basis of the initial MegaBlast results: (i) L1 sequences with ≥ 90% similarity with a known PV (i.e. known PV types) and (ii) L1 sequences with < 90% similarity with any known PV (i.e. putative novel PV types). This subdivision was followed by contig construction and sequences identification using CAP3 and RAxML-EPA, respectively.

Regarding the sequences that share ≥ 90% of identity with known PVs, a total of 8,002,617 reads were generated. The majority were from the genus beta (2,358,670 reads), followed by alpha (1,992,264 reads) and gamma (1,002,061 reads) (Fig. 3A). A substantial proportion of the reads (1,678,061 reads) was assigned to the "unclassified PVs" category, mainly represented by SaPV1 (KX643372.1). The beta-3-1 and beta-3-2 protocols generated a total of 2,588,649 reads (pool 1, Table 2), with a majority (56.6%) of alpha PV sequences, followed by beta HPV sequences (30.8%) (Fig. 3A). The FAP protocol (pool 2) generated 985,675 reads in skin samples, of which 40.8% belonged to the genus beta and 14.5% to

gamma (Table 2; Fig. 3A). The FAPM1 protocol (pool 3) enabled the detection in skin samples of 861,810 reads, comprising alpha (23.3%), beta (13.6%), gamma (14.3%), and mu (7.6%) PV-related sequences (Table 2; Fig. 3A). In oral samples (pool 6), when the same PCR protocol was used, generating 244,587 reads, a different distribution of alpha, beta, and gamma PVs was observed, with 0.1%, 53.6%, and 12.3%, respectively (Table 2; Fig. 3A).

The use of the CUT protocol on skin samples (pool 4) generated 884,923 reads, from the alpha (13%), beta (19.5%), and gamma (23.3%) genera. When the same protocol was used on oral samples (pool 7), generating 78,060 reads, the proportion of alpha (2.1%), beta (11%), and gamma (17.2%) PV-related sequences was different (Table 2; Fig. 3A). The highest proportion of reads (43.4%) generated from this pool corresponded to an unclassified PV (SaPV1).

The FAPM2 protocol (pool 5), used in oral samples, generated 466,004 reads, with a distribution of 9.3% alpha, 39.6% beta, and 32.5% gamma PV-related sequences (Table 2; Fig. 3A).

In addition, products from five PCR protocols (pool 8, Table 2) were pooled and analyzed by NGS. This pool generated 1,892,909 reads, of which 24.5% were representative of beta, 8.8% alpha, and 17.6% gamma PVs. The highest proportion of reads (44.3%) was representative of unclassified PVs (Tables S2).

All the reads correspond to 296 known PV types, including 30 alpha PVs, of which 14 were found in skin samples, 8 in oral samples, and 8 in both tissues. Fifty-four beta HPVs were identified, of which 13 were from the skin, 3 from the oral cavity, and 38 from both tissues. Regarding the genus gamma, 123 known HPVs were identified, of which 70 were isolated from the skin, 8 from the oral cavity, and 45 from both anatomical sites. Three mu HPVs were found (1 in the skin and 2 in both skin and oral samples), and only 1 nu HPV was found (in the skin). Six unclassified PV types were identified, of which 2 were isolated from the skin, 1 from the oral cavity, and 3 from both sites (data not shown).

In addition, 11.3% of the reads (n = 909,308) corresponded to 79 sequences of diverse PVs that do not belong to any of the five PV genera (alpha, beta, gamma, mu, and nu) that contain HPVs; 34 of these 79 sequences were isolated from the skin, 11 from the oral cavity, and 34 from both sites (data not shown).

Regarding the putative novel PVs, we identified 19,032 reads with < 90% similarity with known PVs. The majority of these reads were related to beta (35.6%) and gamma (23.2%) HPV types (Fig. 3B; Table S3). The beta-3-1 and beta-3-2 protocols enabled the identification in pool 1 of 22 reads (26.8%) that are representative of 2 putative new beta-3-related sequences (Fig. 3B; Table S3). In the same pool, 54 reads (65.8%) were assigned to an unclassified PV. However, in the same cluster, a smaller contig was assigned to *Psipapillomavirus* (Table S3). The remaining reads (n = 6, 7.3%) were assigned to *Dyophipapillomavirus 1,* but were matched against HPV115 using the PaVE classification. The FAP protocol enabled the detection in pool 2 of putative new beta (40 reads, 1.2%) and gamma (2228 reads, 69.2%) HPV types. Of the 116 reads that were assigned to unclassified PV using RAxML-EPA, 2 were related to HPV MTS2 (gamma-7) according to the PaVE

classification (Dutta et al., 2017). Finally, 833 reads were identified as *Taupapillomavirus 3,* 4 reads as *Deltapapillomavirus 5,* and 3 reads as *Dyorhopapillomavirus 1* (Table S3).

The FAPM1 protocol enabled the detection in pool 3 of sequences representative of putative new beta (294 reads, 70.2%), gamma (48 reads, 11.5%), delta-2 (52 reads, 12.4%), and lambda-3 (7 reads, 1.7%) PV types. In oral samples, the same protocol enabled the identification of 21 reads, of which 9.5% were found to be related to putative new beta-1 HPV types, 23.8% to *Sigmapapillomavirus 1,* and 66.7% to *Dyoiotapapillomavirus 2* using RAxML-EPA (Fig. 3B; Table S3). No putative new gamma HPV types were identified.

The use of the CUT protocol on skin samples (pool 4) revealed the presence of 2126 reads (72.7%) representative of putative new gamma HPV types. A smaller fraction (4.2%) was representative of putative new beta HPV types. This protocol also revealed the presence of 12 nu and 12 alpha (assigned to species alpha-2 and alpha-3) PV-related reads. RAxML-EPA indicated that 2 reads corresponded to *Lambdapapillomavirus 3,* whereas the same reads got their best initial MegaBlast match against canine papillomavirus 6 (CPV6). Eight other putative non-human PVs were also found (Table S3).

In oral samples (pool 5), when the FAPM2 protocol was used, 6295 reads (99.9%) were representative of putative new beta HPV types, and 0.1% were representative of putative new gamma HPV types (Fig. 3B; Table S3). The CUT protocol (pool 7) enabled the identification in oral samples of only one putative new non-human PV (*Chipapapillomavirus 2*), as well as an unclassified type using RAxML-EPA, but all such reads were assigned to species beta-1 using the PaVE database (Table S3).

Regarding pool 8, only 0.07% and 0.3% of the reads were assigned to beta and mu HPVs, respectively. The remaining reads were related to an unclassified PV (3308 reads), to *Treisdeltapapillomavirus 1* (2713 reads), and to other non-human PV genera (*Treisepsilonpapillomavirus, Treisdeltapapillomavirus*, and *Treiszetapapilloamvirus*; 16 reads).

In summary, all the reads corresponded to 105 putative novel PV types, including 29 beta HPVs, of which 21 were found in skin and 8 in oral samples. Thirty-two gamma HPV types were identified, of which 30 were found in skin and 2 in oral samples. Only 2 putative new alpha HPVs were found in the skin. One mu HPV was found in skin samples. Twenty-four diverse PVs that do not belong to any of the five PV genera that contain HPVs were identified, of which 17 were found in skin and 7 in oral samples. Moreover, 17 unclassified PVs were isolated from skin ($n = 15$) and oral ($n = 2$) samples. However, these reads were found to correspond to beta ($n = 9$) and gamma ($n = 8$) HPVs when the PaVE algorithm was used.

4. Discussion

Since the discovery of the first HPV type four decades ago (Orth et al., 1978), 127 alpha, 93 beta, and 135 gamma HPVs have been described in the Papillomavirus Episteme database (Van Doorslaer et al., 2017). Alpha, beta, and gamma are the most representative genera (Doorbar et al., 2012).

To identify new HPV types, FAP and CUT primers combined with cloning and Sanger sequencing-based strategies have been used successfully in the past (Chouhy et al., 2010; Forslund et al., 1999). However, this approach is quite laborious and time-consuming, and enables the identification of the most represented amplicons only. In particular, this strategy is ineffective in the context of multiple infections. With the advent of new molecular tools (e.g. NGS), the discovery of new HPVs has accelerated over the past few years (Bzhalava et al., 2014; Kocjan et al., 2015). Several studies have shown the capability of NGS in detecting low-copy HPV infections, especially in multiple infections (Arroyo et al., 2013; Barzon et al., 2011; Ekström et al., 2011; Johansson et al., 2013).

Here, we developed a strategy that combines the use of specific or degenerate primers targeting the L1 region of a broad spectrum of HPVs with NGS, for the detection of new HPV types, especially from the genus beta. This strategy incorporates the selective enrichment of PV sequences before NGS is performed. Approximately two thirds of the reads were related to PV sequences. Similar approaches have been reported previously (Arroyo Mühr et al., 2015; Ekström et al., 2013, 2011).

The growing interest in the beta genus arises from evidence that a number of beta HPV types may be involved in pre-malignant and malignant skin lesions (Pfister et al., 2003; Tommasino, 2017). Interestingly, species beta-3 HPV types have been detected in the skin and mucosal epithelia (Forslund et al., 2013; Hampras et al., 2017). Functional studies in *in vitro* and *in vivo* experimental models have highlighted some biological similarities between beta-3 and mucosal HR HPV types. HPV49 shows transforming activity in primary human keratinocytes, and shares some features with HPV16 (Cornet et al., 2012; Viarisio et al., 2016). One of our objectives was to expand the biologically relevant species beta-3, which includes only 4 HPV types, by using beta-3 consensus and degenerate primers.

Combining PCR with novel sets of HPV primers and NGS, we showed the presence of a total of 105 putative new PVs. This procedure also demonstrated the presence of 296 known PV types. Our study showed the presence of a substantial number of beta and gamma HPV types in the oral cavity, which supports the hypothesis of a possible mucosal tropism. However, environmental contamination of the oral cavity cannot be excluded. Furthermore, several other sequences related to unclassified and non-human PVs were identified in skin and oral samples. Environmental contamination may explain the presence of non-human PVs in skin and oral samples. However, cross-species transmission of PVs between animals and humans may also be a consideration (Bravo and Félez-Sánchez, 2015; Gottschling et al., 2011), even though PVs are typically considered to be highly host-restricted (with a few exceptions). Sequences related to bovine PVs have been found in horses and other equids, suggesting interspecies transmission events (Lunardi et al., 2013; Trewby et al., 2014). Other studies also reported cases of cross-species transmission of PVs between bat species (García-Pérez et al., 2014), between rhesus and cynomolgus macaques (Chen et al., 2009), and between humans and cats (Anis et al., 2010; O'Neill et al., 2011); however, additional studies are needed to confirm the latter.

In addition, the notion of “non-human” PV genera needs to be interpreted with caution as they may also include some HPVs. Similarly, alpha and beta genera include few non-human primate PVs (Bernard et al., 2010; Rector and Van Ranst, 2013).

All the results in this study are based on the identification of the sequences using the RAxML-EPA classification. A total of 105 putative new PVs (including 29 beta, 32 gamma, 2 alpha, and 1 mu PVs) were found. In addition, 24 diverse PVs that do not belong to any of the five PV genera that contain HPVs were identified. Interestingly, 17 of the 105 putative new PVs (16.2%) were assigned to taxonomically unclassified PVs. These PVs may not belong to any of the known genera that contain human or animal PVs, and thus may be representative of putative new genera.

The taxonomic assignment performed in this study must be interpreted cautiously, because only small portions of putative new PV genomes have been obtained. In addition, the results obtained using the blastn algorithm refer exclusively to the fraction of the sequence that is aligned by the algorithm. The percentage of similarity indicated by the initial MegaBlast results must also be interpreted with caution, because the definition of novelty for a PV is based on the full L1 ORF length.

In this study, the different protocols were run on different human specimens, and showed different efficacies in detecting putative new PVs, as well as known PVs. The beta-3-1 and beta-3-2 protocols enabled the identification of 4 new beta-3-related sequences in skin samples (using the RAxML-EPA classification), which may potentially expand the beta-3 group to 8 PV types. In vitro experiments are needed to provide insight into the biological properties of these PV types, and to investigate whether these types share biological features with HPV49 (Cornet et al., 2012; Viarisio et al., 2016). The CUT primers enabled the detection of a broad range of PV types in skin and oral samples, including alpha PV types, as previously reported (Chouhy et al., 2010). In contrast, the original FAP protocol was much less likely to identify PVs belonging to the genus alpha. The FAPM1 and FAPM2 protocols enabled the detection of the largest number of putative new PVs in oral samples, whereas the CUT primers enabled the detection of the largest number of putative new PVs in skin samples. Interestingly, the FAPM1 and CUT protocols showed good performance in the detection in skin samples of new PV types that belong to non-human PV genera.

Together, the different protocols enabled the identification of a substantial number ($n = 62$) of putative new beta and gamma HPV types, as well as putative non-human PVs ($n = 24$), in both skin and oral samples.

The gamma HPV types constitute a large group of HPVs that are not yet clearly associated with human disease. However, HPV197, a member of species gamma-24, has recently been detected in human skin cancer specimens (Arroyo Mühr et al., 2015; Grace and Munger, 2017). To date, only 3 HPV types have been classified into the species gamma-24. The use of consensus or degenerate gamma-24 primers might facilitate the discovery of new related PV types, if any exist. Some of the putative new beta or gamma HPV types may also show transforming activity.

In summary, the present study describes a robust strategy based on the use of specific or degenerate primers and NGS technology to detect putative novel PVs. Although the identification of novel PV types or species can only be definitively confirmed by sequencing the whole L1 ORF, initial studies have confirmed the validity of our new protocol as a first step for the isolation and full characterization of novel HPV genomes (e.g. HPV ICB1) (Brancaccio et al., 2017).

The discovery of novel HPV types remains of paramount importance, because new associations between HPV infections and human diseases may be established.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Dr. Karen Muller and Jessica Cox for editing. This study was supported in part by the European Commission project HPV-AHEAD (FP7-HEALTH-2011-282562), by the grant VIRUSCAN R01 (no. R01CA177586-01), and by a grant from “Fondation ARC” (no. PJA 20151203192).

References

Ally MS, Tang JY, Arron ST, 2013. Cutaneous human papillomavirus infection and basal cell carcinoma of the skin. J. Investig. Dermatol 133. 10.1038/jid.2013.46.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. J. Mol. Biol 215, 403–410. 10.1016/S0022-2836C05)80360-2. [PubMed: 2231712]

Andersson K, Waterboer T, Kirnbauer R, Slupetzky K, Iftner T, de Villiers E-M, Forslund O, Pawlita M, Dillner J, 2008. Seroreactivity to cutaneous human papillomaviruses among patients with nonmelanoma skin cancer or benign skin lesions. Cancer Epidemiol. Biomark. Prev 17, 189–195. 10.1158/1055-9965.EPI-07-0405.

Andrews S, 2010. FastQC: a quality control tool for high throughput sequence data.

Anis EA, O’Neill SH, Newkirk KM, Brahmbhatt RA, Abd-Eldaim M, Frank LA, Kania SA, 2010. Molecular characterization of the L1 gene of papillomaviruses in epithelial lesions of cats and comparative analysis with corresponding gene sequences of human and feline papillomaviruses. Am. J. Vet. Res 71, 1457–1461. 10.2460/ajvr.71.12.1457. [PubMed: 21117997]

Arroyo LS, Smelov V, Bzhalava D, Eklund C, Hultin E, Dillner J, 2013. Next generation sequencing for human papillomavirus genotyping. J. Clin. Virol 58, 437–442. 10.1016/j.jcv.2013.07.013. [PubMed: 23932809]

Arroyo Mühr LS, Hultin E, Bzhalava D, Eklund C, Lagheden C, Ekström J, Johansson H, Forslund O, Dillner J, 2015. Human papillomavirus type 197 is commonly present in skin tumors. Int. J. Cancer 136, 2546–2555. 10.1002/ijc.29325. [PubMed: 25388227]

Asgari MM, Kiviat NB, Critchlow CW, Stern JE, Argenyi ZB, Raugi GJ, Berg D, Odland PB, Hawes SE, de Villiers E-M, 2008. Detection of human papillomavirus DNA in cutaneous squamous cell carcinoma among immunocompetent individuals. J. Investig. Dermatol 128, 1409–1417. 10.1038/sj.jid.5701227. [PubMed: 18185530]

Barzon L, Militello V, Lavezzo E, Franchin E, Peta E, Squarzon L, Trevisan M, Pagni S, Dal Bello F, Toppo S, Palù G, 2011. Human papillomavirus genotyping by 454 next generation sequencing technology. J. Clin. Virol 52, 93–97. 10.1016/j.jcv.2011.07.006. [PubMed: 21802982]

Berger SA, Krompass D, Stamatakis A, 2011. Performance, accuracy, and Web server for evolutionary placement of short sequence reads under maximum likelihood. Syst. Biol 60, 291–302. 10.1093/sysbio/syr010. [PubMed: 21436105]

Berger SA, Stamatakis A, 2011. Aligning short reads to reference alignments and trees (https://doi.org/). Bioinformatics 27, 2068–2075. 10.1093/bioinformatics/btr320. [PubMed: 21636595]

Berkhout RJ, Bouwes Bavinck JN, ter Schegget J, 2000. Persistence of human papillomavirus DNA in benign and (pre)malignant skin lesions from renal transplant recipients. J. Clin. Microbiol 38, 2087–2096. [PubMed: 10834958]

Bernard H-U, Burk RD, Chen Z, van Doorslaer K, zur Hausen H, de Villiers E-M, 2010. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology 401, 70–79. 10.1016/j.virol.2010.02.002. [PubMed: 20206957]

Bouvard V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L, Cogliano V, WHO International Agency for Research on Cancer Monograph Working Group, 2009. A review of human carcinogens–Part B: biological agents. Lancet Oncol. 10, 321–322. [PubMed: 19350698]

Bouwes Bavinck JN, Neale RE, Abeni D, Euvrard S, Green AC, Harwood CA, de Koning MNC, Naldi L, Nindl I, Pawlita M, Pfister H, Proby CM, Quint WGV, ter Schegget J, Waterboer T, Weissenborn S, Feltkamp MCW, EPI-HPV-UV-CA group, 2010. Multicenter study of the association between betapapillomavirus infection and cutaneous squamous cell carcinoma. Cancer Res. 70, 9777–9786. 10.1158/0008-5472.CAN-10-0352. [PubMed: 21098702]

Brancaccio RN, Robitaille A, Dutta S, Rollison DE, Fischer N, Grundhoff A, Tommasino M, Gheit T, 2017. Complete genome sequence of a novel human gammapapillomavirus isolated from skin. Genome Announc. 5. 10.1128/genomeA.00833-17.

Bravo IG, Félez-Sánchez M, 2015. Papillomaviruses: viral evolution, cancer and evolutionary medicine. Evol. Med. Public Health 2015, 32–51. 10.1093/emph/eov003. [PubMed: 25634317]

Bzhalava D, Eklund C, Dillner J, 2015. International standardization and classification of human papillomavirus types. Virology 476, 341–344. 10.1016/j.virol.2014.12.028. [PubMed: 25577151]

Bzhalava D, Guan P, Franceschi S, Dillner J, Clifford G, 2013. A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types. Virol. Spec. Issue.: Papillomavirus Epistem. 445, 224–231. 10.1016/j.virol.2013.07.015.

Bzhalava D, Mühr LSA, Lagheden C, Ekström J, Forslund O, Dillner J, Hultin E, 2014. Deep sequencing extends the diversity of human papillomaviruses in human skin. Sci. Rep 4, 5807. 10.1038/srep05807. [PubMed: 25055967]

Casabonne D, Michael KM, Waterboer T, Pawlita M, Forslund O, Burk RD, Travis RC, Key TJ, Newton R, 2007. A prospective pilot study of antibodies against human papillomaviruses and cutaneous squamous cell carcinoma nested in the Oxford component of the European Prospective investigation into cancer and Nutrition. Int. J. Cancer 121, 1862–1868. 10.1002/ijc.22885. [PubMed: 17565742]

Chen Z, van Doorslaer K, DeSalle R, Wood CE, Kaplan JR, Wagner JD, Burk RD, 2009. Genomic diversity and interspecies host infection of alpha12 Macaca fascicularis papillomaviruses (MfPVs). Virology 393, 304–310. 10.1016/j.virol.2009.07.012. [PubMed: 19716580]

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD, 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31, 3497–3500. [PubMed: 12824352]

Chouhy D, Gorosito M, Sánchez A, Serra EC, Bergero A, Fernandez Bussy R, Giri AA, 2010. New generic primer system targeting mucosal/genital and cutaneous human papillomaviruses leads to the characterization of HPV 115, a novel Beta-papillomavirus species 3. Virology 397, 205–216. 10.1016/j.virol.2009.11.020. [PubMed: 19948351]

Cornet I, Bouvard V, Campo MS, Thomas M, Banks L, Gissmann L, Lamartine J, Sylla BS, Accardi R, Tommasino M, 2012. Comparative analysis of transforming properties of E6 and E7 from different beta human papillomavirus types. J. Virol 86, 2366–2370. 10.1128/JVI.06579-11. [PubMed: 22171257]

de Jong-Tieben LM, Berkhout RJ, Smits HL, Bouwes Bavinck JN, Vermeer BJ, van der Woude FJ, ter Schegget J, 1995. High frequency of detection of epidermodysplasia verruciformis-associated human papillomavirus DNA in biopsies from malignant and premalignant skin lesions from renal transplant recipients. J. Investig. Dermatol 105, 367–371. [PubMed: 7665914]

de Villiers E-M, Fauquet C, Broker TR, Bernard H-U, zur Hausen H, 2004. Classification of papillomaviruses. Virology 324, 17–27. 10.1016/j.virol.2004.03.033. [PubMed: 15183049]

Donà MG, Gheit T, Vescio MF, Latini A, Moretto D, Benevolo M, Cristaudo A, Tommasino M, Giuliani M, 2016. Incidence, clearance and duration of cutaneous beta and gamma human papillomavirus anal infection. J. Infect. 73, 380–383. 10.1016/j.jinf.2016.07.006. [PubMed: 27427205]

Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, Stanley MA, 2012. The biology and life-cycle of human papillomaviruses. Vaccine 30 (Suppl 5), F55–70. 10.1016/j.vaccine.2012.06.083. [PubMed: 23199966]

Dutta S, Robitaille A, Olivier M, Rollison DE, Tommasino M, Gheit T, 2017. Genome sequence of a novel human gammapapillomavirus isolated from skin. Genome Announc. 5. 10.1128/genomeA.00439-17.

Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. 10.1093/nar/gkh340. [PubMed: 15034147]

Ekström J, Bzhalava D, Svenback D, Forslund O, Dillner J, 2011. High throughput sequencing reveals diversity of Human Papillomaviruses in cutaneous lesions. Int. J. Cancer 129, 2643–2650. 10.1002/ijc.26204. [PubMed: 21630257]

Ekström J, Mühr LSA, Bzhalava D, Söderlund-Strand A, Hultin E, Nordin P, Stenquist B, Paoli J, Forslund O, Dillner J, 2013. Diversity of human papillomaviruses in skin lesions. Virology 447, 300–311. 10.1016/j.virol.2013.09.010. [PubMed: 24210127]

Forslund O, Antonsson A, Nordin P, Stenquist B, Hansson BG, 1999. 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), 2437–2443. 10.1099/0022-1317-80-9-2437. [PubMed: 10501499]

Forslund O, Johansson H, Madsen KG, Kofoed K, 2013. The nasal mucosa contains a large spectrum of human papillomavirus types from the betapapillomavirus and gammapapillomavirus genera. J. Infect. Dis 208, 1335–1341. 10.1093/infdis/jit326. [PubMed: 23878325]

García-Pérez R, Ibáñez C, Godínez JM, Aréchiga N, Garin I, Pérez-Suárez G, de Paz O, Juste J, Echevarría JE, Bravo IG, 2014. Novel papillomaviruses in free-ranging Iberian bats: no virus-host co-evolution, no strict host specificity, and hints for recombination. Genome Biol. Evol 6, 94–104. 10.1093/gbe/evt211. [PubMed: 24391150]

Giuliano A, Lazcano E, Villa L, Flores R, Salmeron J, Lee J-H, Papenfuss M, Abrahamsen M, Baggio M, Silva R, Quiterio M, 2009. Circumcision and sexual behavior: factors independently associated with human papillomavirus (HPV) detection among men in The HIM study. Int. J. Cancer J. Int. Cancer 124, 1251–1257. 10.1002/ijc.24097.

Giuliano AR, Lazcano-Ponce E, Villa LL, Flores R, Salmeron J, Lee J-H, Papenfuss MR, Abrahamsen M, Jolles E, Nielson CM, Baggio ML, Silva R, Quiterio M, 2008a. The human papillomavirus infection in men study: human papillomavirus prevalence and type distribution among men residing in Brazil, Mexico, and the United States. Cancer Epidemiol. Biomark. Prev 17, 2036–2043. 10.1158/1055-9965.EPI-08-0151.

Giuliano AR, Lee J-H, Fulp W, Villa LL, Lazcano E, Papenfuss MR, Abrahamsen M, Salmeron J, Anic GM, Rollison DE, Smith D, 2011. Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study. Lancet 377, 932–940. 10.1016/S0140-6736(10)62342-2. [PubMed: 21367446]

Giuliano AR, Tortolero-Luna G, Ferrer E, Burchell AN, de Sanjose S, Kjaer SK, Muñoz N, Schiffman M, Bosch FX, 2008b. Epidemiology of human papillomavirus infection in men, in cancers other than cervical and in benign conditions. Vaccine 26, K17–K28. 10.1016/j.vaccine.2008.06.021. [PubMed: 18847554]

Goon P, Sonnex C, Jani P, Stanley M, Sudhoff H, 2008. Recurrent respiratory papillomatosis: an overview of current thinking and treatment. Eur. Arch. Otorhinolaryngol 265, 147–151. 10.1007/s00405-007-0546-z. [PubMed: 18046565]

Gottschling M, Göker M, Stamatakis A, Bininda-Emonds ORP, Nindl I, Bravo IG, 2011. Quantifying the phylodynamic forces driving papillomavirus evolution. Mol. Biol. Evol 28, 2101–2113. 10.1093/molbev/msr030. [PubMed: 21285031]

Grace M, Munger K, 2017. Proteomic analysis of the gamma human papillomavirus type 197 E6 and E7 associated cellular proteins. Virology 500, 71–81. 10.1016/j.virol.2016.10.010. [PubMed: 27771561]

Haedicke J, Iftner T, 2013. Human papillomaviruses and cancer. Radiother. Oncol 108, 397–402. 10.1016/j.radonc.2013.06.004. [PubMed: 23830197]

Hampras SS, Giuliano AR, Lin H-Y, Fisher KJ, Abrahamsen ME, McKay-Chopin S, Gheit T, Tommasino M, Rollison DE, 2015. Natural history of polyomaviruses in men: the HPV infection in men (HIM) study. J. Infect. Dis 211,1437–1446. 10.1093/infdis/jiu626. [PubMed: 25387582]

Hampras SS, Giuliano AR, Lin H-Y, Fisher KJ, Abrahamsen ME, Sirak BA, Iannacone MR, Gheit T, Tommasino M, Rollison DE, 2014. Natural history of cutaneous human papillomavirus (HPV) infection in men: the HIM study. PLoS One 9. 10.1371/journal.pone.0104843.

Hampras SS, Rollison DE, Giuliano AR, McKay-Chopin S, Minoni L, Sereday K, Gheit T, Tommasino M, 2017. Prevalence and concordance of cutaneous beta human papillomavirus infection at mucosal and cutaneous sites. J. Infect. Dis 216, 92–96. 10.1093/infdis/jix245. [PubMed: 28549147]

Harwood CA, Surentheran T, McGregor JM, Spink PJ, Leigh IM, Breuer J, Proby CM, 2000. Human papillomavirus infection and non-melanoma skin cancer in immunosuppressed and immunocompetent individuals. J. Med. Virol 61, 289–297. [PubMed: 10861635]

Huang X, Madan A, 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877. [PubMed: 10508846]

Iannacone MR, Gheit T, Pfister H, Giuliano AR, Messina JL, Fenske NA, Cherpelis BS, Sondak VK, Roetzheim RG, Silling S, Pawlita M, Tommasino M, Rollison DE, 2014. Case-control study of genus-beta human papillomaviruses in plucked eyebrow hairs and cutaneous squamous cell carcinoma. Int. J. Cancer 134, 2231–2244. 10.1002/ijc.28552. [PubMed: 24136717]

Iannacone MR, Gheit T, Waterboer T, Giuliano AR, Messina JL, Fenske NA, Cherpelis BS, Sondak VK, Roetzheim RG, Michael KM, Tommasino M, Pawlita M, Rollison DE, 2012. Case-control study of cutaneous human papillomaviruses in squamous cell carcinoma of the skin. Cancer Epidemiol. Biomark. 21, 1303–1313. 10.1158/1055-9965.EPI-12-0032.

Iftner A, Klug SJ, Garbe C, Blum A, Stancu A, Wilczynski SP, Iftner T, 2003. The prevalence of human papillomavirus genotypes in nonmelanoma skin cancers of nonimmunosuppressed individuals identifies high-risk genital types as possible risk factors. Cancer Res. 63, 7515–7519. [PubMed: 14612553]

Johansson H, Bzhalava D, Ekström J, Hultin E, Dillner J, Forslund O, 2013. Metagenomic sequencing of “HPV-negative” condylomas detects novel putative HPV types. Virology 440, 1–7. 10.1016/j.virol.2013.01.023. [PubMed: 23522725]

Käller M, Ewels P, 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics.

Karagas MR, Nelson HH, Sehr P, Waterboer T, Stukel TA, Andrew A, Green AC, Bavinck JNB, Perry A, Spencer S, Rees JR, Mott LA, Pawlita M, 2006. Human papillomavirus infection and incidence of squamous cell and basal cell carcinomas of the skin. J. Natl. Cancer Inst 98, 389–395. 10.1093/jnci/djj092. [PubMed: 16537831]

Kocjan BJ, Bzhalava D, Forslund O, Dillner J, Poljak M, 2015. Molecular methods for identification and characterization of novel papillomaviruses. Clin. Microbiol. Infect 21, 808–816. 10.1016/j.cmi.2015.05.011. [PubMed: 26003284]

Krueger F, 2015. Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files.

Kumar S, Stecher G, Tamura K, 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol 33, 1870–1874. 10.1093/molbev/msw054. [PubMed: 27004904]

Lunardi M, de Alcântara BK, Otonel RAA, Rodrigues WB, Alfieri AF, Alfieri AA, 2013. Bovine papillomavirus type 13 DNA in equine sarcoids. J. Clin. Microbiol 51, 2167–2171. 10.1128/JCM.00371-13. [PubMed: 23637294]

Mahé F, Rognes T, 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ

Matsen FA, Hoffman NG, Gallagher A, Stamatakis A, 2012. A format for phylogenetic placements. PLoS One 7, e31009. 10.1371/journal.pone.0031009. [PubMed: 22383988]

Myers G, Baker GMC, Münger K, Sverdrup F, McBride A & Bernard HU, 1997. Alignments. In Human Papillomaviruses 1997. HPV Sequence Database II-L1–23–73.

Myers G, Baker GMC, Münger K, Sverdrup F, McBride A &Bernard HU, 1996. Alignments. In Human Papillomaviruses 1996. HPV Sequence Database II-L1–1–67.

Nei Kumar, 2000. Molecular Evolution and Phylogenetics. Oxford University Press.

Nunes EM, Sudenga SL, Gheit T, Tommasino M, Baggio ML, Ferreira S, Galan L, Silva RC, Pierce Campbell CM, Lazcano-Ponce E, Giuliano AR, Villa LL, Sichero L, 2016. Diversity of beta-papillomavirus at anogenital and oral anatomic sites of men: the HIM Study. Virology 495, 33–41. 10.1016/j.virol.2016.04.031. [PubMed: 27161202]

Nyitray AG, Carvalho da Silva RJ, Baggio ML, Lu B, Smith D, Abrahamsen M, Papenfuss M, Villa LL, Lazcano-Ponce E, Giuliano AR, 2011. Age-specific prevalence of and risk factors for anal human papillomavirus (HPV) among men who have sex with women and men who have sex with men: the HPV in men (HIM) study. J. Infect. Dis 203, 49–57. 10.1093/infdis/jiq021. [PubMed: 21148496]

Ondov BD, Bergman NH, Phillippy AM, 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinform 12, 385. 10.1186/1471-2105-12-385.

O'Neill SH, Newkirk KM, Anis EA, Brahmbhatt R, Frank LA, Kania SA, 2011. Detection of human papillomavirus DNA in feline premalignant and invasive squamous cell carcinoma. Vet. Dermatol 22, 68–74. 10.1111/j.1365-3164.2010.00912.x. [PubMed: 20609206]

Orth G, 2006. Genetics of epidermodysplasia verruciformis: insights into host defense against papillomaviruses. Semin. Immunol 18, 362–374. 10.1016/j.smim.2006.07.008. [PubMed: 17011789]

Orth G, Jablonska S, Favre M, Croissant O, Jarzabek-Chorzelska M, Rzesa G, 1978. Characterization of two types of human papillomaviruses in lesions of epidermodysplasia verruciformis. Proc. Natl. Acad. Sci. USA 75, 1537–1541. [PubMed: 206906]

Pfister H, 2003. Chapter 8: human papillomavirus and skin cancer. J. Natl. Cancer Inst. Monogr 52–56. [PubMed: 12807946]

Pfister H, Fuchs PG, Majewski S, Jablonska S, Pniewska I, Malejczyk M, 2003. High prevalence of epidermodysplasia verruciformis-associated human papillomavirus DNA in actinic keratoses of the immunocompetent population. Arch. Dermatol. Res 295, 273–279. 10.1007/s00403-003-0435-2. [PubMed: 14618345]

Pierce Campbell CM, Gheit T, Tommasino M, Lin H-Y, Torres BN, Messina JL, Stoler MH, Rollison DE, Sirak BA, Abrahamsen M, Carvalho da Silva RJ, Sichero L, Villa LL, Lazcano-Ponce E, Giuliano AR, 2016. Cutaneous beta human papillomaviruses and the development of male external genital lesions: a case-control study nested within the HIM Study. Virology 497, 314–322. 10.1016/j.virol.2016.08.002. [PubMed: 27518539]

Pierce Campbell CM, Messina JL, Stoler MH, Jukic DM, Tommasino M, Gheit T, Rollison DE, Sichero L, Sirak BA, Ingles DJ, Abrahamsen M, Lu B, Villa LL, Lazcano-Ponce E, Giuliano AR, 2013. Cutaneous human papillomavirus types detected on the surface of male external genital lesions: a case series within the HPV Infection in Men Study. J. Clin. Virol 58, 652–659. 10.1016/j.jcv.2013.10.011. [PubMed: 24210970]

Rector A, Van Ranst M, 2013. Animal papillomaviruses. Virology 445, 213–223. 10.1016/j.virol.2013.05.007. [PubMed: 23711385]

Smelov V, Hanisch R, McKay-Chopin S, Sokolova O, Eklund C, Komyakov B, Gheit T, Tommasino M, 2017. Prevalence of cutaneous beta and gamma human papillomaviruses in the anal canal of men who have sex with women. Papillomavirus Res. 3, 66–72. 10.1016/j.pvr.2017.02.002. [PubMed: 28720458]

Stamatakis A, 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. 10.1093/bioinformatics/btu033. [PubMed: 24451623]

Tommasino M, 2017. The biology of beta human papillomaviruses. Virus Res. 231, 128–138. 10.1016/j.virusres.2016.11.013. [PubMed: 27856220]

Tommasino M, 2014. The human papillomavirus family and its role in carcinogenesis. Semin. Cancer Biol 26, 13–21. 10.1016/j.semcancer.2013.11.002. [PubMed: 24316445]

Trewby H, Ayele G, Borzacchiello G, Brandt S, Campo MS, Del Fava C, Marais J, Leonardi L, Vanselow B, Biek R, Nasir L, 2014. Analysis of the long control region of bovine papillomavirus type 1 associated with sarcoids in equine hosts indicates multiple cross-species transmission events

and phylogeographical structure. J. Gen. Virol 95, 2748–2756. 10.1099/vir.0.066589-0. [PubMed: 25185436]

Van Doorslaer K, Li Z, Xirasagar S, Maes P, Kaminsky D, Liou D, Sun Q, Kaur R, Huyen Y, McBride AA, 2017. The Papillomavirus Episteme: a major update to the papillomavirus sequence database. Nucleic Acids Res. 45, D499–D506. 10.1093/nar/gkw879. [PubMed: 28053164]

Van Doorslaer K, Tan Q, Xirasagar S, Bandaru S, Gopalan V, Mohamoud Y, Huyen Y, McBride AA, 2013. The papillomavirus Episteme: a central resource for papillomavirus sequence data and analysis. Nucleic Acids Res. 41, D571–D578. 10.1093/nar/gks984. [PubMed: 23093593]

Viarisio D, Müller-Decker K, Zanna P, Kloz U, Aengeneyndt B, Accardi R, Flechtenmacher C, Gissmann L, Tommasino M, 2016. Novel ß-HPV49 transgenic mouse model of upper digestive tract cancer. Cancer Res. 76, 4216–4225. 10.1158/0008-5472.CAN-16-0370. [PubMed: 27216183]

Waterboer T, Abeni D, Sampogna F, Rother A, Masini C, Sehr P, Michael KM, Pawlita M, 2008. Serological association of beta and gamma human papillomaviruses with squamous cell carcinoma of the skin. Br. J. Dermatol 159, 457–459. 10.1111/j.1365-2133.2008.08621.x. [PubMed: 18503604]

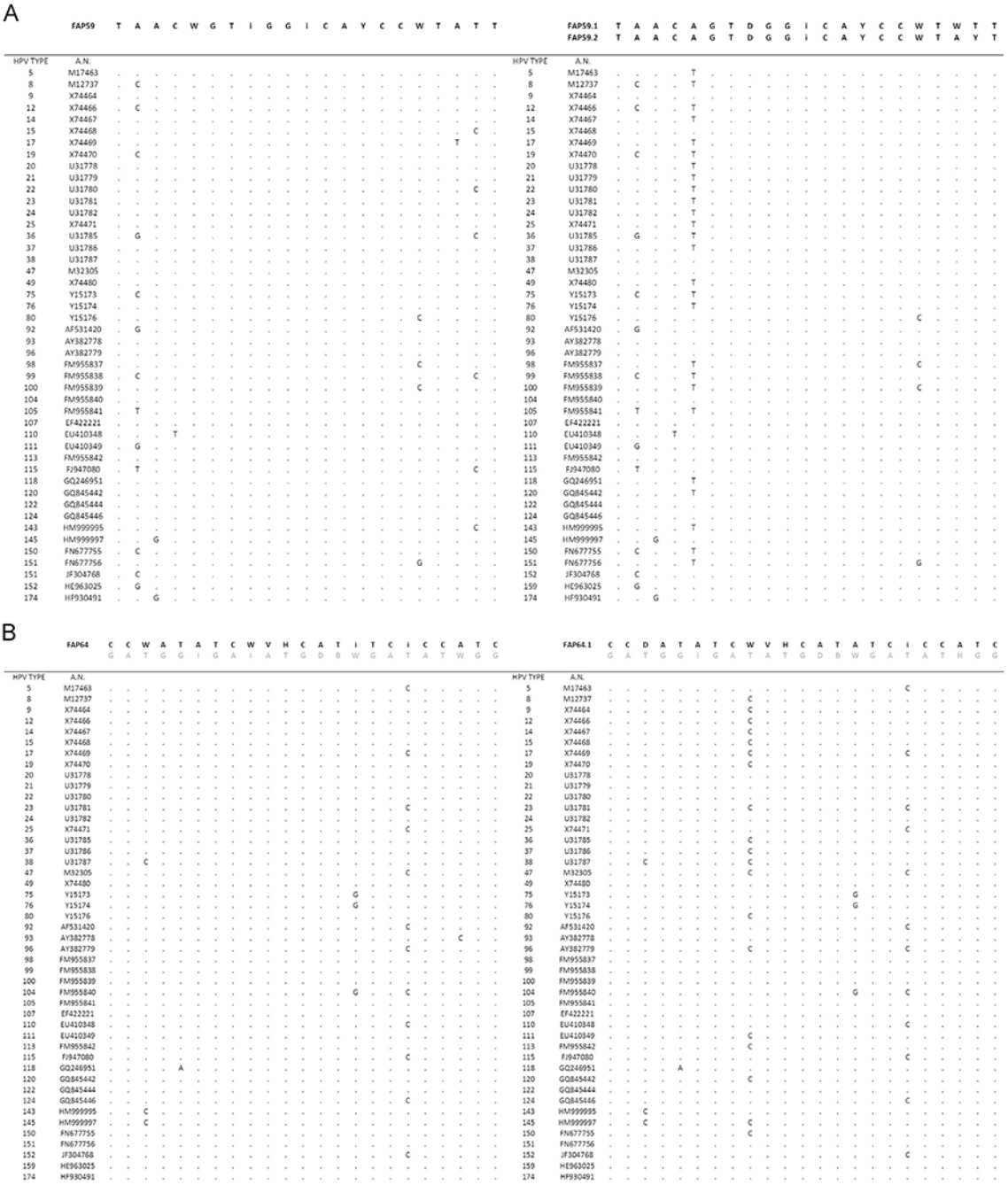

Fig. 1.

Construction of improved FAP degenerate primers by alignment of 46 beta HPV type L1 regions with MUSCLE 3.8.

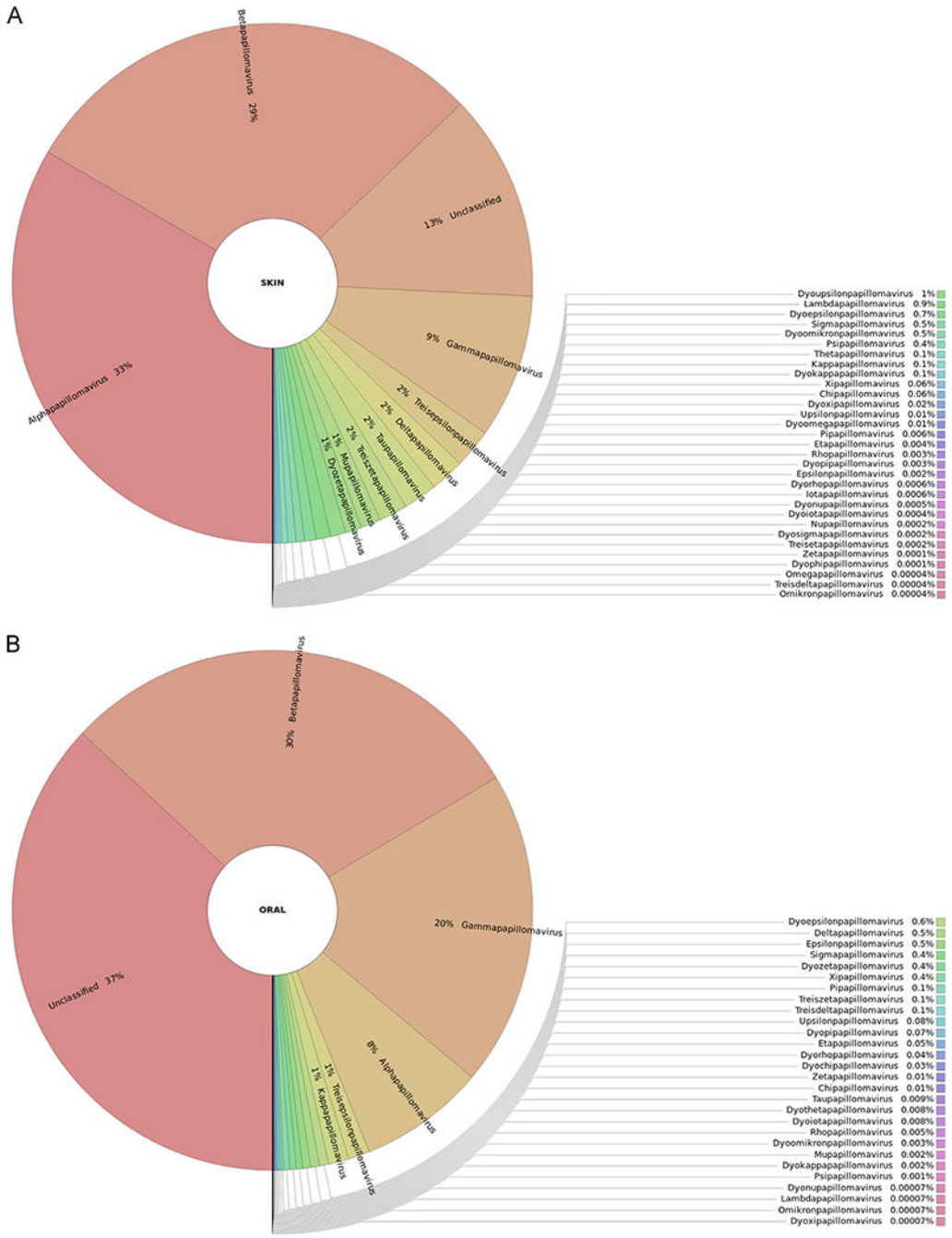

Fig. 2.

Graphical representation of the unnormalized abundance of PV genera and species in terms of number of reads: (A) skin samples, (B) oral samples.

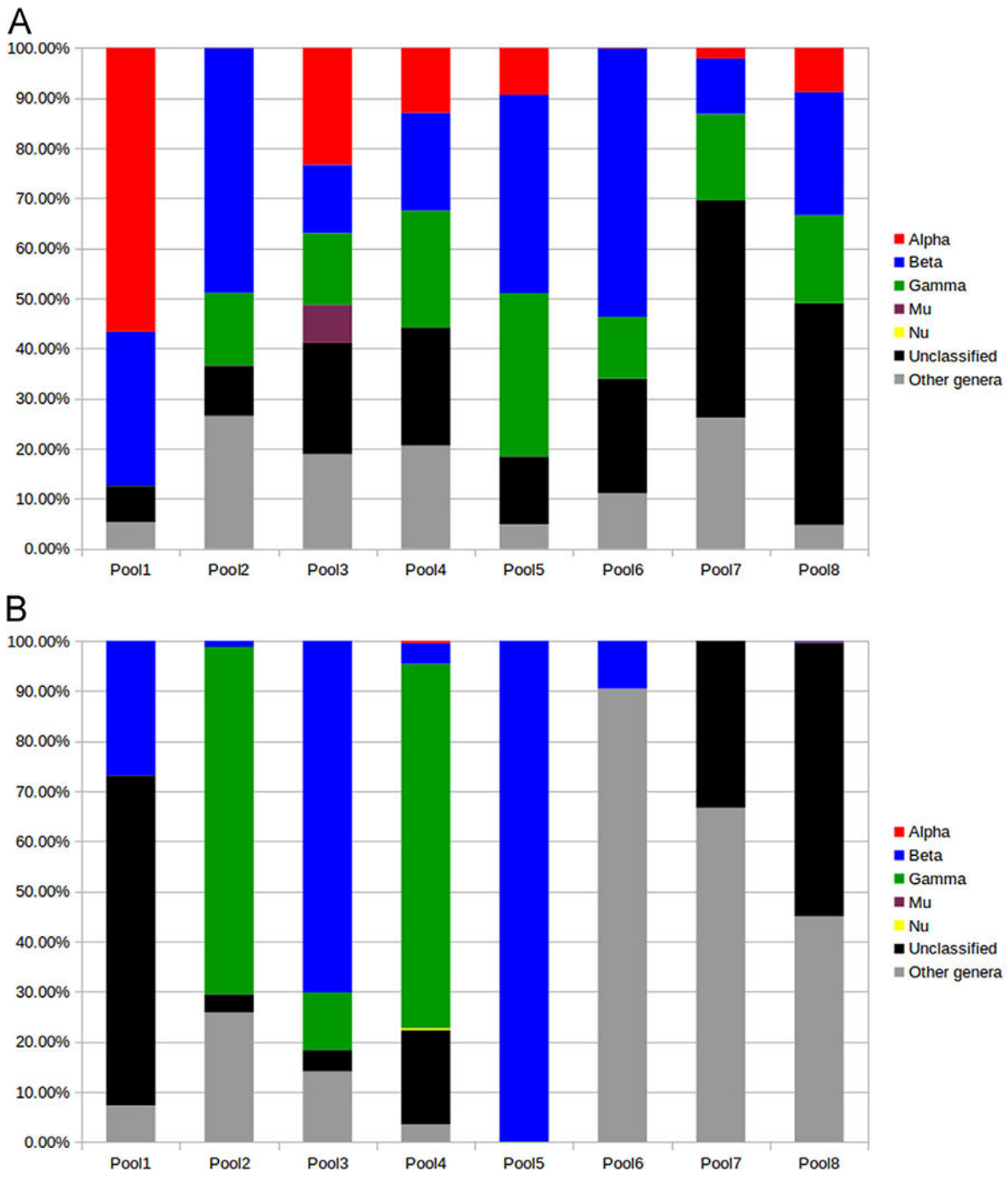

Fig. 3.

(A) Distribution of the known PVs detected in the different NGS pools, in terms of percentage of reads within each pool; (B) distribution of the putative new PVs detected in the different NGS pools, in terms of percentage of reads within each pool (RAxML-EPA).

Table 1

Sequences of the oligonucleotides and composition of the different protocols. i = inosine; W = A or T; D = A or G or T; K = T or G; Y = C or T; M = A or C; R = A or G; V = A or C or G; H = A or C or T.

Primer mix	Primer sequence (5–3')
Beta-3-1	
B3L1FW3	AGGACATCCATACTTTGAGGTTCGAG
B3L1FW4	TAGGACATCCATATTTTGATGTGAGAG
B3L1FW5	GATGTTAGAGACACTGGAGATTCAACA
B3L1FW6	GATGTTAGAGACACTGGGGATTCAACA
B3L1FW7	GATGTTAGAGACACTGTGGATCAAACA
B3L1RW	ATAATAGTATTTCTTAATTCTAATGGAGG
B3L1RW4	ATAACTGAATTGATTAATTCTAATGGAGG
B3L1RW5	ATAACTGTATTTACTAATTCTAAAGGTGG
B3L1RW6	TACAGTATTTACCAGTTCCAAAGGTGG
B3L1RW7	ATTACAGTATTAACTAATTCTAAAGGTGG
B3L1RW8	ATTACAGTATTTACTAATTCTAAAGGTGG
Beta-3-2	
B3L1FW1	GTAGGACATCCATAYTTTGAKGTKiGAG
B3L1FW2	TTGATGTTAGAGACACTGiDGATYMAACA
B3L1RW1	ATAAiWGWATTKYTTAATTCTAATGGAGG
B3L1RW2	ATTACAGTATTiACKARTTCYAAAGGTGG
CUT	
CUT1Fw	TRCCiGAYCCiAATAARTTTG
CUT1AFw	TRCCiGAYCCiAACAGRTTTG
CUT1BFw	TRCCiGAYCCiAATAGRTTTG
CUT1CFw	TRCCiGAYCCiAACAARTTTG
CUT1BRv	ARGAYGGiGAYATGGTiGA
FAP	
FAP59	TAACWGTiGGiCAYCCWTATT
FAP64	CCWATATCWVHCATiTCiCCATC
FAPM1	
FAP59.1	TAACAGTDGGiCAYCCWTWTT
FAP59.2	TAACAGTDGGiCAYCCWTAYT
FAP64.1	CCDATATCWVHCATATCiCCATC
FAP59	TAACWGTiGGiCAYCCWTATT
FAP64	CCWATATCWVHCATiTCiCCATC
FAPM2	
FAP59.2	TAACAGTDGGiCAYCCWTAYT
FAP64.1	CCDATATCWVHCATATCiCCATC

Table 2

Description of the PCR protocols and NGS pools.

PCR pools	PCR protocols	Specimens	N	NGS pools
1	Beta-3-1	Skin swab	41	1
2	Beta-3-2		9	
3	FAP		52	2
4	FAPM1		54	3
5	CUT		57	4
6	FAPM2	Oral gargle	43	5
7	FAPM1		56	6
8	CUT		55	7
9	Beta-3-1		9	8
10	Beta-3-2		4	
11	FAP		11	
12	FAPM1		11	
13	FAPM2		12	