Journal of General Virology (2013), **94**, 000–000 DOI 10.1099/vir.0.055137-0

Short Communication

Analysis of the genetic diversity and phylogenetic relationships of putative human papillomavirus types

Diego Chouhy,[1,2] Elisa M. Bolatti,[2] Germán R. Pérez[1,2] and Adriana A. Giri[1,2]

Correspondence
Adriana A. Giri
giri@ibr-conicet.gov.ar

[1]Area Virología, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina

[2]Instituto de Biología Molecular y Celular de Rosario (CONICET), Rosario, Argentina

Received 16 May 2013
Accepted 24 August 2013

More than 170 human papillomavirus (HPV) types have been completely sequenced, curated and divided into five genera: *Alphapapillomavirus*, *Betapapillomavirus*, *Gammapapillomavirus*, *Mupapillomavirus* and *Nupapillomavirus*. With the application of PCR methods, hundreds of putative novel HPV types have been identified as PCR amplicons in mucosa and skin. However, at present there are no studies reporting a systematic search of the currently known L1 amplicons and their phylogenetic relationships. This survey revealed the existence of at least 202 different putative HPV types that are pending for full-genome characterization: five alphapapillomaviruses, 37 betapapillomaviruses, 159 gammapapillomaviruses and one mupapillomavirus. All potential viruses of the genera *Alphapapillomavirus* and *Betapapillomavirus* were grouped in the defined species, while 59 putative gammapapillomaviruses types were segregated in 21 unidentified putative species. These data highlight the need for progress in the identification of additional taxa of the family *Papillomaviridae* in order to elucidate the diversity, evolution and medical implications of these viruses.

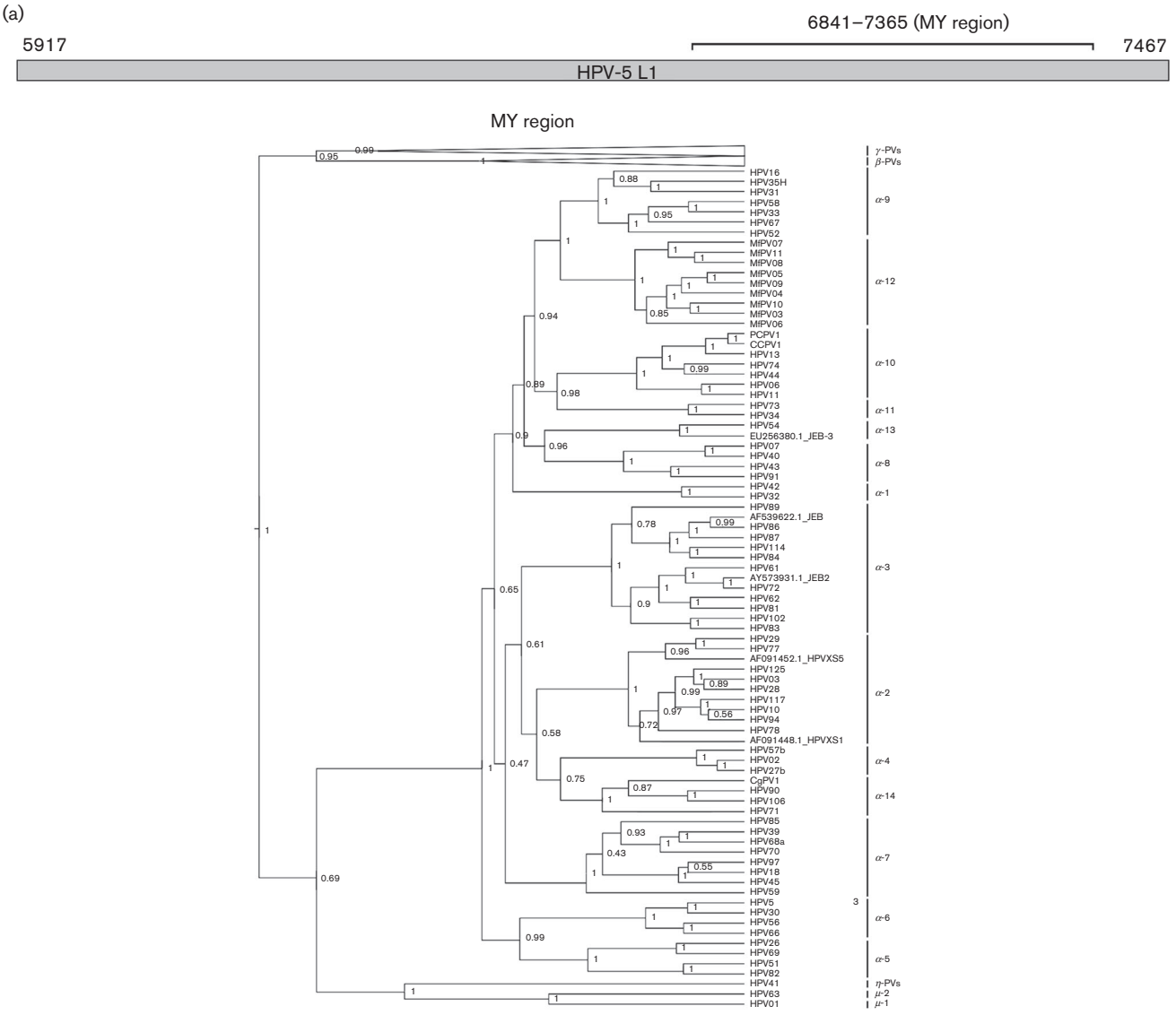
Papillomaviruses (PVs) are small non-enveloped DNA tumour viruses with a circular genome of nearly 8 kb. PVs infect the epithelia of vertebrates and are host specific (Bernard *et al.*, 2010). By convention, designation of a novel PV type requires the genome to be cloned and curated by the Papillomavirus Reference Center and show less than 90 % identity in the L1 ORF with respect to any known PV type, while PV types belonging to new species within a genus share 60–70 % nucleotide identity with PV types within this genus (de Villiers *et al.*, 2004).

Presently, 170 human PV (HPV) types have been officially designated and completely sequenced (de Villiers, 2013; http://www.hpvcenter.se/html/refclones.html), and divided into five genera according to their L1 ORF phylogenetic relationships: *Alphapapillomavirus*, *Betapapillomavirus*, *Gammapapillomavirus*, *Mupapillomavirus* and *Nupapillomavirus*. There is diversity in the pathology of the HPV types across the genera and species, in particular in relation to the epithelium infected and the oncogenic potential of the viral type. The genus *Alphapapillomavirus* is heterogeneous, containing: (i) the high-risk mucosal HPV types in species α-7 and α-9; (ii) low-risk mucocutaneous genital types in species α-10; and (iii) viruses grouped in the species α-4, which are most frequently associated with common skin warts (Bernard *et al.*, 2010; de Villiers *et al.*, 2004).

Other HPV-containing genera are less heterogeneous. The betapapillomavirus (*β*-PV) types cause flat lesions in epidermodysplasia verruciformis patients, but evident HPV lesions are rare in immune-competent individuals (Pfister *et al.*, 2003). The mupapillomavirus (μ-PV) and some gammapapillomavirus (*γ*-PV) types cause proliferative cutaneous lesions in humans, although the recently identified *γ*-PV types are not associated with known lesions. In addition, *γ*-PV and μ-PV types differ from each other in that the E5 ORF is missing from members of the genus *Gammapapillomavirus*, as it is in members of the genus *Betapapillomavirus*. However, multiple *β*-PV and *γ*-PV types have been identified in the oral [e.g. HPV-120 and HPV-145 (*β*-2); 124 (*β*-1); HPV-121 (*γ*-10); HPV-134 (*γ*-7)] (Bottalico *et al.*, 2011; Kocjan *et al.*, 2011) and cervical [e.g. HPV-109 (*γ*-7), HPV-112 (*γ*-8); HPV-101, HPV-103 and HPV-108 (*γ*-6)] mucosa (Bernard *et al.*, 2010). These findings raise new questions about the anatomical tissue tropisms, the evolution of HPVs and the epidemiological associations of HPV with oral and skin neoplasia.

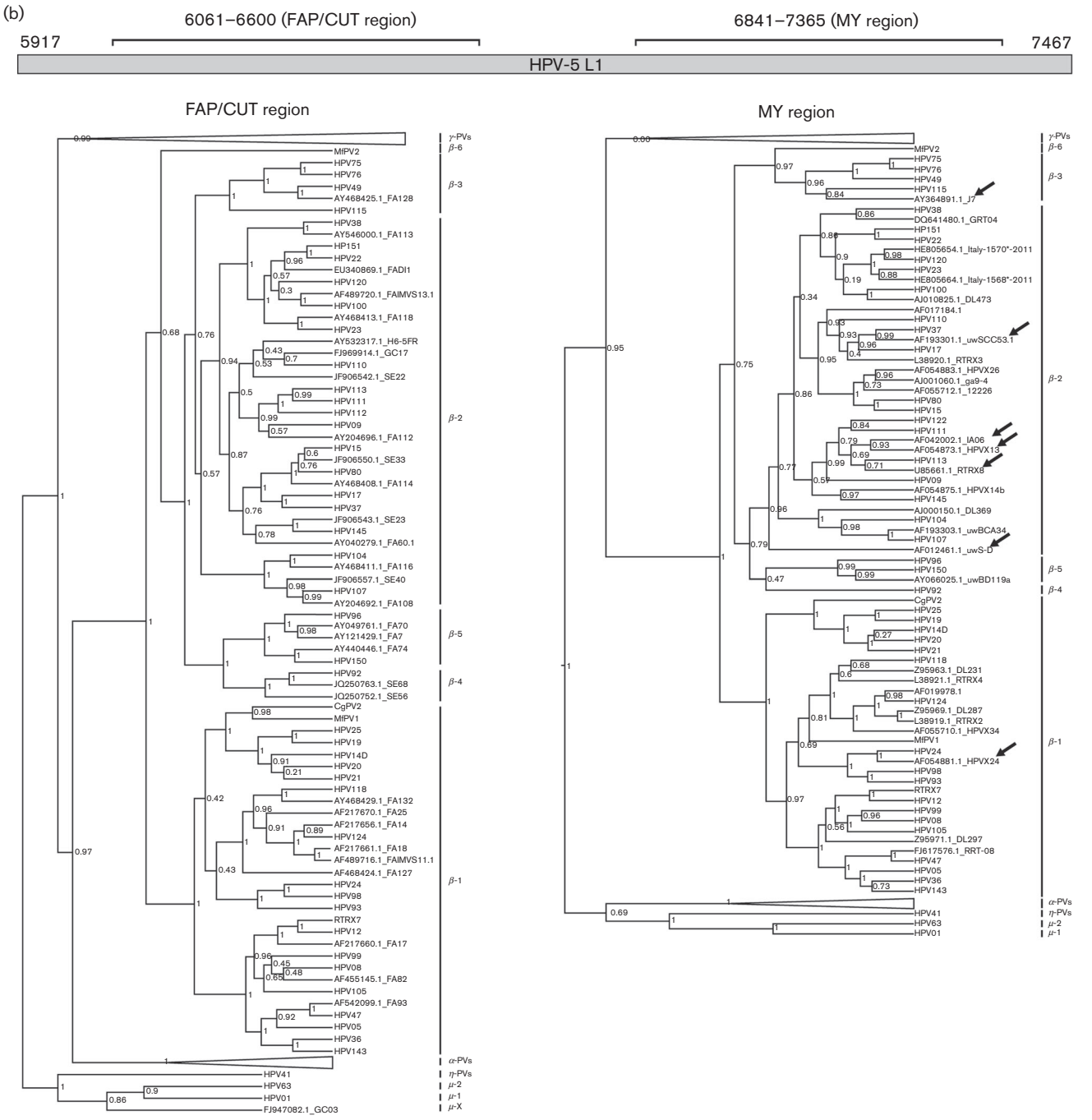
To date, 66 α-PVs, 45 *β*-PVs, 54 *γ*-PVs, two μ-PVs and one nupapillomavirus (*η*-PV) have been isolated from humans

One supplementary figure is available with the online version of this paper.



(de Villiers, 2013), but it is assumed that a higher number exists in nature. With the development of different PCR methods for cutaneous and/or mucosal screening, hundreds of putative HPV types have been identified in the form of PCR amplicons of 80–450 bp in mucosa (Menzo *et al.*, 2001; Peyton & Wheeler, 1994) and skin lesions (Asgari *et al.*, 2008; Berkhout *et al.*, 2000; Chouhy *et al.*, 2010; Forslund *et al.*, 2003, 2007; Harwood *et al.*, 2000, 2004; Shamanin *et al.*, 1994, 1996), as well as in normal skin (Antonsson *et al.*, 2000; Chen *et al.*, 2008; Chouhy *et al.*, 2010, 2013; Hazard *et al.*, 2007b). Previously, phylogenetic analyses showed that many HPV types/putative HPV types, particularly those belonging to the genus *Gammapapillomavirus*, are found segregated outside the current species (Chouhy *et al.*, 2010, 2013; Forslund, 2007), indicating the anticipation of additional species groups. There are no present studies reporting a systematic search of the currently known L1 amplicons or its putative phylogenetic relationships. Also, the identification of new putative HPV species may help to direct the attention for generic or specific primer designs in order to speed up the complete characterization of HPV genomes.

The aim of this study was to update the knowledge of the HPV putative types identified to date by the use of different primer systems targeting the L1 gene, and to determine their phylogenetic associations. With that purpose, a systematic search was done in the GenBank database using the term 'papillomaviridae L1'. A total of 5509 sequences were retrieved that included complete genome PV sequences or partial PV sequences of different PV genes. In order to keep subgenomic sequences corresponding to HPV putative types, complete genome sequences, non-human PV sequences and partial sequences with truncated protein sequences were removed. Putative HPV types and novel putative species within a genus were defined according to the current criteria based on the nucleotide identities in the L1 fragment sequences (de Villiers *et al.*,



2004). Next, those sequences with more than 90 % nucleotide identity with a previously known HPV type were removed to exclude subtypes and variants of characterized HPV types. The final number of sequences corresponding to putative HPV types was 223, with lengths ranging from 90 to 480 nt (<200 nt=8; 200–400 nt=73; >400 nt=142). These putative HPV sequences split into two major regions: (i) 190 were in the Forslund–Antonsson primers (FAP) (Forslund *et al.*, 1999)/CUT (Chouhy *et al.*, 2010) primers region (nt 6061–6600 in the HPV-5 genome; GenBank accession no. NC_001531); and (ii) 33 were in the MY (Manos *et al.*, 1989) primers region (nt 6841–7365 in the HPV-5 genome).

Subsequently, the taxonomic validity of the FAP/CUT and MY sequence regions was investigated. Pairwise comparisons of the complete L1 sequences, the FAP/CUT and the MY sequence regions of 166 different characterized PV types (151 HPVs and 15 non-human PVs) were compared with respect to each closest relative type. This analysis revealed a mean difference between known defined HPV types and closely related partial fragments of 1.6 % (range,

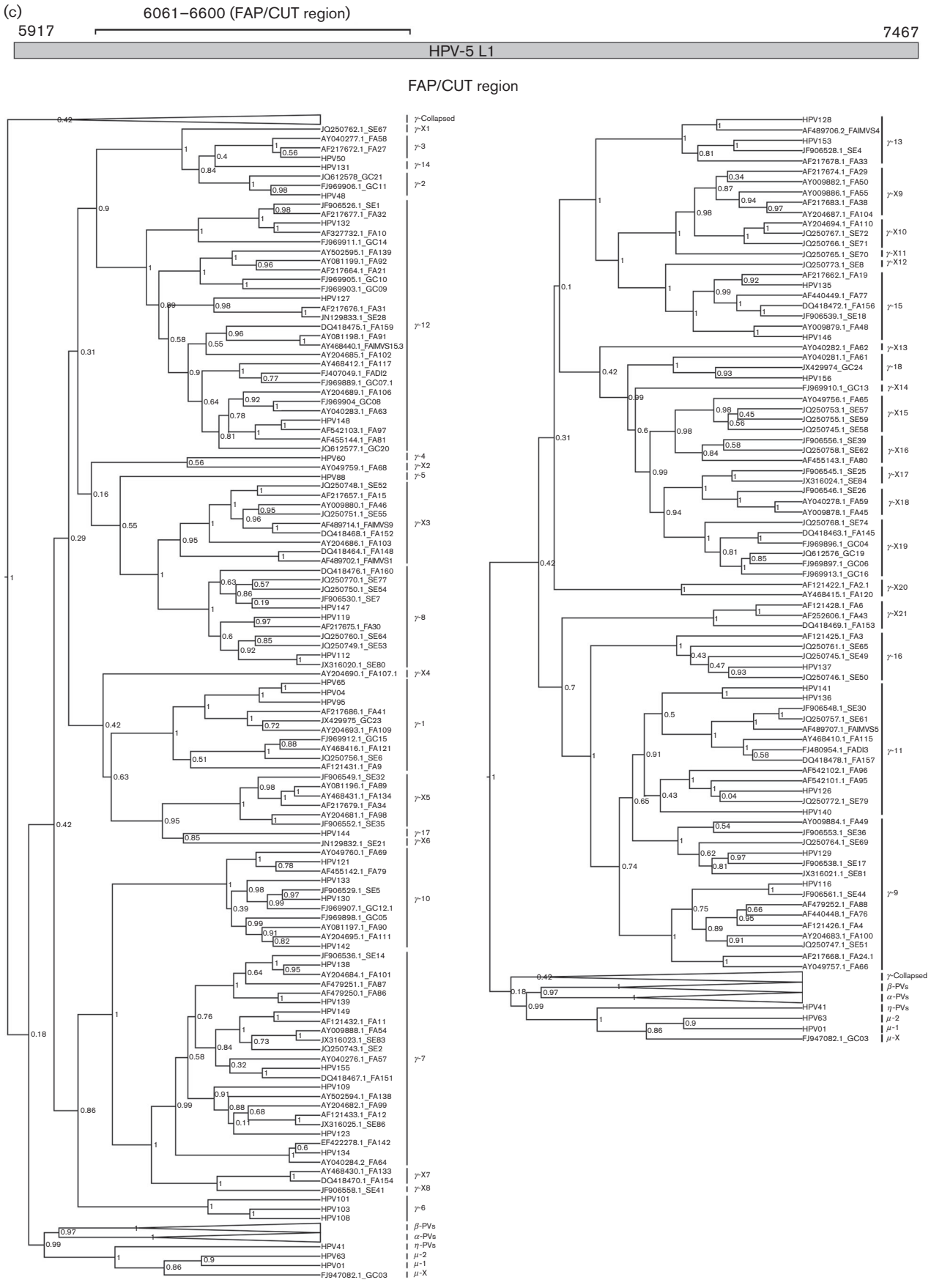

(c)
6061–6600 (FAP/CUT region)
5917
7467
HPV-5 L1
FAP/CUT region
γ-Collapsed
JQ250762.1_SE67
AY040277.1_FA58
AF217672.1_FA27
HPV50
HPV131
JQ612578_GC21
FJ969906.1_GC11
HPV48
γ-X1
γ-3
γ-14
γ-2
JF906526.1_SE1
AF217677.1_FA32
HPV132
AF327732.1_FA10
FJ969911.1_GC14
AY502595.1_FA139
AY081199.1_FA92
AF217664.1_FA21
FJ969905.1_GC10
FJ969903.1_GC09
HPV127
AF217676.1_FA31
JN129833.1_SE28
DQ418475.1_FA159
AY081198.1_FA91
AY468440.1_FAIMVS15.3
AY204685.1_FA102
AY468412.1_FA117
FJ407049.1_FADI2
FJ969889.1_GC07.1
AY204689.1_FA106
FJ969904_GC08
AY040283.1_FA63
HPV148
AF542103.1_FA97
AF455144.1_FA81
JQ612577.1_GC20
γ-12
HPV60
AY049759.1_FA68
HPV88
γ-4
γ-X2
γ-5
JQ250748.1_SE52
AF217657.1_FA15
AY009880.1_FA46
JQ250751.1_SE55
AF489714.1_FAIMVS9
DQ418468.1_FA152
AY204686.1_FA103
DQ418464.1_FA148
AF489702.1_FAIMVS1
γ-X3
DQ418476.1_FA160
JQ250770.1_SE77
JQ250750.1_SE54
JF906530.1_SE7
HPV147
HPV119
AF217675.1_FA30
JQ250760.1_SE64
JQ250749.1_SE53
HPV112
JX316020.1_SE80
γ-8
AY204690.1_FA107.1
γ-X4
HPV65
HPV04
HPV95
AF217686.1_FA41
JX429975_GC23
AY204693.1_FA109
FJ969912.1_GC15
AY468416.1_FA121
JQ250756.1_SE6
AF121431.1_FA9
γ-1
JF906549.1_SE32
AY081196.1_FA89
AY468431.1_FA134
AF217679.1_FA34
AY204681.1_FA98
JF906552.1_SE35
γ-X5
HPV144
JN129832.1_SE21
γ-17
γ-X6
AY049760.1_FA69
HPV121
AF455142.1_FA79
HPV133
JF906529.1_SE5
HPV130
FJ969907.1_GC12.1
FJ969898.1_GC05
AY081197.1_FA90
AY204695.1_FA111
HPV142
γ-10
JF906536.1_SE14
HPV138
AY204684.1_FA101
AF479251.1_FA87
AF479250.1_FA86
HPV139
HPV149
AF121432.1_FA11
AY009888.1_FA54
JX316023.1_SE83
JQ250743.1_SE2
AY040276.1_FA57
HPV155
DQ418467.1_FA151
HPV109
AY502594.1_FA138
AY204682.1_FA99
AF121433.1_FA12
JX316025.1_SE86
HPV123
EF422278.1_FA142
HPV134
AY040284.2_FA64
γ-7
AY468430.1_FA133
DQ418470.1_FA154
γ-X7
JF906558.1_SE41
γ-X8
HPV101
HPV103
HPV108
γ-6
β-PVs
α-PVs
η-PVs
μ-2
μ-1
μ-X
HPV41
HPV63
HPV01
FJ947082.1_GC03
HPV128
AF489706.2_FAIMVS4
HPV153
JF906528.1_SE4
AF217678.1_FA33
γ-13
AF217674.1_FA29
AY009882.1_FA50
AY009886.1_FA55
AF217683.1_FA38
AY204687.1_FA104
γ-X9
AY204694.1_FA110
JQ250767.1_SE72
JQ250766.1_SE71
γ-X10
JQ250765.1_SE70
γ-X11
JQ250773.1_SE8
γ-X12
AF217662.1_FA19
HPV135
AF440449.1_FA77
DQ418472.1_FA156
JF906539.1_SE18
AY009879.1_FA48
HPV146
γ-15
AY040282.1_FA62
AY040281.1_FA61
JX429974_GC24
HPV156
γ-X13
γ-18
FJ969910.1_GC13
γ-X14
AY049756.1_FA65
JQ250753.1_SE57
JQ250755.1_SE59
JQ250745.1_SE58
γ-X15
JF906556.1_SE39
JQ250758.1_SE62
AF455143.1_FA80
γ-X16
JF906545.1_SE25
JX316024.1_SE84
JF906546.1_SE26
AY040278.1_FA59
AY009878.1_FA45
γ-X17
γ-X18
JQ250768.1_SE74
DQ418463.1_FA145
FJ969896.1_GC04
JQ612576_GC19
FJ969897.1_GC06
FJ969913.1_GC16
γ-X19
AF121422.1_FA2.1
AY468415.1_FA120
γ-X20
AF121428.1_FA6
AF252606.1_FA43
DQ418469.1_FA153
γ-X21
AF121425.1_FA3
JQ250761.1_SE65
JQ250745.1_SE49
HPV137
JQ250746.1_SE50
γ-16
HPV141
HPV136
JF906548.1_SE30
JQ250757.1_SE61
AF489707.1_FAIMVS5
AY468410.1_FA115
FJ480954.1_FADI3
DQ418478.1_FA157
AF542102.1_FA96
AF542101.1_FA95
HPV126
JQ250772.1_SE79
HPV140
γ-11
AY009884.1_FA49
JF906553.1_SE36
JQ250764.1_SE69
HPV129
JF906538.1_SE17
JX316021.1_SE81
HPV116
JF906561.1_SE44
AF479252.1_FA88
AF440448.1_FA76
AF121426.1_FA4
AY204683.1_FA100
JQ250747.1_SE51
γ-9
AF217668.1_FA24.1
AY049757.1_FA66
γ-Collapsed
β-PVs
α-PVs
η-PVs
μ-2
μ-1
μ-X
HPV41
HPV63
HPV01
FJ947082.1_GC03

Fig. 1. Phylogenetic trees of 190 putative HPV types identified in the FAP/CUT region and 33 putative HPV types found in the MY region, and 166 characterized PV types. (a) Genus *Alphapapillomavirus*. (b) Genus *Betapapillomavirus*. (c) Genus *Gammapapillomavirus*: the left part of the tree is indicated as 'γ-Collapsed' in the right part of the figure and vice versa. Protein sequence-derived nucleotide multiple alignments were performed with MEGA5 (Tamura *et al.*, 2011), and phylogenetic relationships were inferred by Bayesian analysis using BEAST 1.7.2 (Drummond *et al.*, 2012). To do so, Markov Chain Monte Carlo simulations were performed on 10^7 generations, sampling one state every 1000 generations, with a burn-in of 10 %. Evolutionary substitution model for each run was set as the rtREV + Γ + I. Node values represent BPP. Each putative HPV type is noted by its GenBank accession number followed by the isolate name. For reasons of clarity, the other genera different from those under analysis, were collapsed in each subtree, except for those containing few taxa (μ-PVs and η-PVs). Only species containing putative HPV types in either subgenomic region are shown. Arrows indicate putative HPV types in the MY region that are potentially different from those in the FAP/CUT region. γ-X*n*, additional putative species inferred in this work.

0–4.7 %) nucleotide identity in the FAP/CUT region, and of 1.6 % (range, 0–5 %) nucleotide identity in the MY region with respect to the L1 ORF of each closest relative type across the 166 viral types. These values were in agreement with the mean differences of nucleotide identities between the L1 ORF and the FA subgenomic sequences obtained by other authors (Forslund, 2007; Hazard *et al.*, 2007a). Therefore, a putative HPV type showing 90.0 ± 1.6 % nucleotide identity with its closest relative in either FAP/CUT or MY regions may indicate a novel type worth being completely cloned and sequenced.

Comparative analysis of trees based on the L1 ORF and in the subgenomic sequences under analysis demonstrated largely congruent tree topologies and high Bayesian posterior probability (BPP) values (Fig. S1, available in JGV Online), indicating that the FAP/CUT and the MY sequence regions were useful for phylogenetic analyses. In most cases each type had the same closest relative PV type in both partial regions when compared with the tree constructed from alignments of the L1 ORF. The only exception was HPV-126 (γ-11), which was grouped in the species γ-9 in both FAP/CUT and MY regions phylogenies. Even so, the phylogenies obtained with both L1 subgenomic regions were suitable to group the types, and therefore the putative HPV types, in the defined species to which they belonged.

The phylogenetic analyses of the FAP/CUT region from 190 putative HPV types and 166 reference PVs classified 30 as β-PVs (Fig. 1b, left), 159 as γ-PVs (Fig. 1c) and one as μ-PV (Fig. 1). On the other hand, the phylogenetic analyses of the MY region from 33 putative HPV types and 166 reference PVs classified five strains as α-PVs (Fig. 1a) and 28 as β-PVs (Fig. 1b, right). Clearly, the primers used for HPV identification have different specificities, those designed in the MY region being more specific for the genera *Alphapapillomavirus* and *Betapapillomavirus*, while those targeting the FAP/CUT region are more specific for genera *Betapapillomavirus* and *Gammapapillomavirus*. In fact, putative HPV types of the genus *Alphapapillomavirus* were identified only in the MY region, those grouped in the genus *Gammapapillomavirus* were found exclusively in the FAP/CUT region, while strains belonging to the genus *Betapapillomavirus* were found in both subgenomic regions.

To go further in the identification of the different putative HPV types existing in nature so far, and considering that both subgenomic regions do not overlap, we selected all 190 putative HPV types found in the FAP/CUT region (30 β-PVs, 159 γ-PVs and one μ-PV), five putative α-PV types identified in the MY region and seven β-PV strains located in the MY region presenting different closely related HPV types in the FAP/CUT region (Fig. 1b, see arrows). This criterion was applied to avoid considering the same strain as two different putative HPV types when appearing in both subgenomic regions (i.e. β-PVs, Fig. 1b). Therefore, this survey indicates that out of 223 subgenomic sequences identified, at least 202 may correspond to different putative HPV types (five α-PVs, 37 β-PVs, 159 γ-PVs and one μ-PV). Table 1 summarizes all HPV subgenomic sequences identified in this report, including the 21 potentially redundant strains, which are depicted in bold.

According to their phylogenetic associations, all putative HPV types identified in the genera *Alphapapillomavirus* (Fig. 1a) and *Betapapillomavirus* (Fig. 1b) were segregated inside the defined species. In contrast, only 98 out of the 159 putative types found in the genus *Gammapapillomavirus* were segregated in the defined species (γ-1–γ-17) (Fig. 1c, Table 1) (de Villiers, 2013). Moreover, two putative HPV types belonged to a recently identified species (γ-18) (Chouhy *et al.*, 2013), and 59 were segregated in 21 unidentified putative species (named γ-X1–γ-X21) (Fig. 1c, Table 1). In addition, the strain GC03 (Fig. 1) may probably define a new putative species within the genus *Mupapillomavirus* (69.8 and 68.2 % nucleotide identities with HPV-1a and HPV-63, respectively). Comparative analysis of trees generated by the maximum-likelihood method (RaxML program, evolutionary substitution model set as rtREV + Γ + I with fast bootstrap of 1000 replicates) revealed almost the same phylogenetic associations as those obtained by Bayesian analysis (data not shown), further supporting the overall relationships found and the existence of additional undefined putative species within the genera γ-PV and μ-PV as well. Interestingly, those species of the genera *Betapapillomavirus* and *Gammapapillomavirus* containing viruses that showed incongruences in the phylogenetic analysis of the FAP/CUT and MY regions (Fig. S1: β-1, β-2, γ-7, γ-8, γ-10, γ-11, γ-12) had the greatest number of putative HPV types within both subgenomic regions (FAP/

Table 1. Main features of the 223 putative HPV types identified in this work

GenBank accession numbers/strains shown in bold indicate the 21 potentially redundant putative HPV types. γ-X*n*, additional putative species inferred in this work.

GenBank accession no./ strain	Species	Epithelium/ source†	GenBank accession no./ strain	Species	Epithelium/source†	GenBank accession no./ strain	Species	Epithelium/ source†
AF091448.1/HPVXS1	α-2	S/NR	AF121433.1/FA12	γ-7	S/BSL	JQ612577/GC20	γ-12	S/HS IC
AF091452.1/HPVXS5	α-2	S/NR	AF479250.1/FA86	γ-7	S/HS IC	AF217678.1/FA33	γ-13	S/HS IS
AF539622.1/JEB	α-3	NR/NR	AF479251.1/FA87	γ-7	Environment	AF489706.2/FAIMVS4	γ-13	S/HS IC
AY573931.1/JEB2	α-3	NR/NR	AY009888.1/FA54	γ-7	S/HS IC	JF906528.1/SE4	γ-13	S/SCC-AK
EU256380.1/JEB-3	α-13	M/ASCUS	AY040276.1/FA57	γ-7	S/HS IC	AF217662.1/FA19	γ-15	S/HS IS
AF019978.1	β-1	S/EV Scar Ca	AY040284.2/FA64	γ-7	S/HS IC	AF440449.1/FA77	γ-15	S/HS IC
AF054881.1/HPVX24	β-1	S/NR	AY204682.1/FA99	γ-7	NR/NR	AY009879.1/FA48	γ-15	S/HS IC
AF055710.1/HPVX34	β-1	S/BSL	AY204684.1/FA101	γ-7	NR/NR	DQ418472.1/FA156	γ-15	S/HS IC
AF217656.1/FA14	β-1	S/HS IS	AY502594.1/FA138	γ-7	S/AK	JF906539.1/SE18	γ-15	S/SCC-AK
AF217660.1/FA17	β-1	S/HS IS	DQ418467.1/FA151	γ-7	S/HS IC	AF121425.1/FA3	γ-16	S/HS IC
AF217661.1/FA18	β-1	S/HS IS	EF422278.1/FA142	γ-7	S/SCC	JQ250745.1/SE49	γ-16	NR/NR
AF217670.1/FA25	β-1	S/HS IS	JF906536.1/SE14	γ-7	S/SCC-AK	JQ250746.1/SE50	γ-16	NR/NR
AF455145.1/FA82	β-1	S/HS IC	JQ250743.1/SE2	γ-7	S/SCC-AK	JQ250761.1/SE65	γ-16	NR/NR
AF489716.1/FAIMVS11.1	β-1	S/SCC (PL)	JX316023.1/SE83	γ-7	NR/NR	AY040281.1/FA61	γ-18	S/HS IC
AF542099.1/FA93	β-1	S/HS IC	JX316025.1/SE86	γ-7	NR/NR	JX429974/GC24	γ-18	S/HS IC
AY468424.1/FA127	β-1	S/SeK	AF217675.1/FA30	γ-8	S/HS IS	JQ250762.1/SE67	γ-X1	NR/NR
AY468429.2/FA132	β-1	S/HS IC	DQ418476.1/FA160	γ-8	NR/NR	AY049759.1/FA68	γ-X2	S/HS IC
FJ617576.1/RRT-08	β-1	S/EV	JF906530.1/SE7	γ-8	S/BSL	AF217657.1/FA15	γ-X3	S/HS IS
L38919.1/RTRX2	β-1	S/SCC	JQ250749.1/SE53	γ-8	NR/NR	AF489702.1/FAIMVS1	γ-X3	S/SoK
L38921.1/RTRX4	β-1	S/SCC	JQ250750.1/SE54	γ-8	NR/NR	AF489714.1/FAIMVS9	γ-X3	S/SCC
Z95963.1/DL231	β-1	M/SCC	JQ250760.1/SE64	γ-8	NR/NR	AY009880.1/FA46	γ-X3	S/HS IC
Z95969.1/DL287	β-1	S/SCC	JQ250770.1/SE77	γ-8	NR/NR	AY204686.1/FA103	γ-X3	NR/NR
Z95971.1/DL297	β-1	S/melanoma	JX316020.1/SE80	γ-8	NR/NR	DQ418464.1/FA148	γ-X3	S/HS IC
AF012461.1/uwS-D	β-2	M/intestine	AF121426.1/FA4	γ-9	S/HS IC	DQ418468.1/FA152	γ-X3	S/HS IC
AF017184.1	β-2	S/SCC	AF217668.1/FA24.1	γ-9	S/HS IC	JQ250748.1/SE52	γ-X3	NR/NR
AF042002.1/IA06	β-2	M/oral cavity	AF440448.1/FA76	γ-9	S/HS IC	JQ250751.1/SE55	γ-X3	NR/NR
AF054873.1/HPVX13‡	β-2	S/AK	AF479252.1/FA88	γ-9	S/HS IC	AY204690.1/FA107.1	γ-X4	NR/NR
AF054875.1/HPVX14b‡	β-2	S/AK	AY009884.1/FA49	γ-9	S/HS IC	AF217679.1/FA34	γ-X5	S/HS IS
AF054883.1/HPVX26	β-2	S/NR	AY049757.1/FA66	γ-9	S/HS IC	AY081196.1/FA89	γ-X5	S/HS IC
AF055712.1/12226	β-2	NR/NR	AY204683.1/FA100	γ-9	NR/NR	AY204681.1/FA98	γ-X5	NR/NR
AF193301.1/uwSCC53.1	β-2	S/SCC	JF906538.1/SE17	γ-9	S/SCC-AK	AY468431.1/FA134	γ-X5	S/HS IC
AF193303.1/uwBCA34	β-2	S/Bowen Ca	JF906553.1/SE36	γ-9	S/BSL	JF906549.1/SE32	γ-X5	S/SCC-AK
AF489720.1/FAIMVS13.1	β-2	S/BCC (PL)	JF906561.1/SE44	γ-9	S/SCC-AK	JF906552.1/SE35	γ-X5	S/BSL
AJ000150.1/DL369	β-2	M/SCC	JQ250747.1/SE51	γ-9	NR/NR	JN129832.1/SE21	γ-X6	S/SCC-AK
AJ001060.1/ga9-4	β-2	S/HS IC	JQ250764.1/SE69	γ-9	NR/NR	AY468430.1/FA133	γ-X7	S/HS IC
AJ010825.1/DL473	β-2	S/SCC	JX316021.1/SE81	γ-9	NR/NR	DQ418470.1/FA154	γ-X7	S/HS IC
AY040279.1/FA60.1	β-2	S/HS IC	AF455142.1/FA79	γ-10	S/HS IC	JF906558.1/SE41	γ-X8	S/SCC-AK
AY204692.1/FA108	β-2	NR/NR	AY049760.1/FA69‡	γ-10	S/HS IC	AF217674.1/FA29	γ-X9	S/HS IS
AY204696.1/FA112	β-2	NR/NR	AY081197.1/FA90	γ-10	S/HS IC	AF217683.1/FA38	γ-X9	S/HS IC
AY468408.1/FA114	β-2	S/BCC	AY204695.1/FA111	γ-10	NR/NR	AY009882.1/FA50	γ-X9	S/HS IC

Table 1. cont.

GenBank accession no./ strain	Species	Epithelium/ source†	GenBank accession no./ strain	Species	Epithelium/source†	GenBank accession no./ strain	Species	Epithelium/ source†
AY468411.1/FA116	β-2	S/BCC	FJ969898.1/GC05	γ-10	S/AK	AY009886.1/FA55	γ-X9	S/HS IC
AY468413.1/FA118	β-2	S/SCC	FJ969907.1/GC12.1	γ-10	S/HS IC	AY204687.1/FA104	γ-X9	NR/NR
AY532317.1/H6-5FR	β-2	S/HS IC	JF906529.1/SE5	γ-10	S/BSL	AY204694.1/FA110	γ-X10	NR/NR
AY546000.1/FA113	β-2	NR/NR	AF489707.1/FAIMVS5	γ-11	S/BCC	JQ250766.1/SE71	γ-X10	NR/NR
DQ641480.1/GRT04	β-2	S/VW	AF542101.1/FA95	γ-11	S/HS IC	JQ250767.1/SE72	γ-X10	NR/NR
EU340869.1/FADI1	β-2	S/HS IC	AF542102.1/FA96	γ-11	S/HS IC	JQ250765.1/SE70	γ-X11	NR/NR
FJ969914.1/GC17	β-2	S/HS IC	AY468410.1/FA115	γ-11	S/HS IC	JQ250773.1/SE8	γ-X12	S/BSL
HE805654.1/Italy 1570*-2011	β-2	Wastewater	DQ418473.1/FA157	γ-11	S/HS IC	AY040282.1/FA62	γ-X13	S/HS IC
HE805664.1/Italy 1568*-2011	β-2	Wastewater	FJ480954.1/FADI3‡	γ-11	S/HS IC	FJ969910.1/GC13	γ-X14	S/VW
JF906542.1/SE22	β-2	S/SCC-AK	JF906548.1/SE30	γ-11	S/SCC-AK	AY049756.1/FA65	γ-X15	S/HS IC
JF906543.1/SE23	β-2	S/SCC-AK	JQ250757.1/SE61‡	γ-11	NR/NR	JQ250753.1/SE57	γ-X15	NR/NR
JF906550.1/SE33	β-2	S/SCC-AK	JQ250772.1/SE79	γ-11	NR/NR	JQ250754.1/SE58	γ-X15	NR/NR
JF906557.1/SE40	β-2	S/SCC-AK	AF217664.1/FA21	γ-12	S/HS IC	JQ250755.1/SE59	γ-X15	NR/NR
L38920.1/RTRX3‡	β-2	S/SCC	AF217676.1/FA31	γ-12	S/HS IS	AF455143.1/FA80	γ-X16	S/HS IC
U85661.1/RTRX8	β-2	S/SCC	AF217677.1/FA32	γ-12	S/HS IS	JF906556.1/SE39	γ-X16	S/SCC-AK
AY364891.1/J7	β-3	M/NR	AF327732.1/FA10	γ-12	NR/NR	JQ250758.1/SE62	γ-X16	NR/NR
AY468425.1/FA128	β-3	S/SCC	AF455144.1/FA81	γ-12	S/HS IC	JF906545.1/SE25	γ-X17	S/BSL
JQ250752.1/SE56	β-4	NR/NR	AF542103.1/FA97	γ-12	S/HS IC	JX316024.1/SE84	γ-X17	NR/NR
JQ250763.1/SE68	β-4	NR/NR	AY040283.1/FA63	γ-12	S/HS IC	AY009878.1/FA45	γ-X18	Environment
AF121429.1/FA7	β-5	S/HS IC	AY081198.1/FA91	γ-12	S/HS IC	AY040278.1/FA59	γ-X18	S/HS IC
AF440446.1/FA74	β-5	S/HS IC	AY081199.1/FA92	γ-12	S/HS IC	JF906546.1/SE26	γ-X18	S/SCC-AK
AY049761.1/FA70	β-5	S/HS IC	AY204685.1/FA102	γ-12	NR/NR	DQ418463.1/FA145	γ-X19	S/HS IC
AY066025.1/uwBD119a	β-5	S/BD	AY204689.1/FA106	γ-12	NR/NR	FJ969896.1/GC04	γ-X19	S/HS IC
AF121431.1/FA9	γ-1	S/HS IC	AY468412.1/FA117	γ-12	S/BCC	FJ969897.1/GC06	γ-X19	S/AK (PL)
AF217686.1/FA41	γ-1	S/HS IS	AY468440.1/FAIMVS15.3	γ-12	S/BCC	FJ969913.1/GC16	γ-X19	S/HS IC
AY204693.1/FA109	γ-1	NR/NR	AY502595.1/FA139	γ-12	S/SeK	JQ250768.1/SE74	γ-X19	NR/NR
AY468416.1/FA121	γ-1	S/BCC	DQ418475.1/FA159	γ-12	NR/NR	JQ612576/GC19	γ-X19	S/HS IC
FJ969912.1/GC15‡	γ-1	S/HS IC	FJ407049.1/FADI2	γ-12	S/HS IC	AF121422.1/FA2.1	γ-X20	S/HS IC
JQ250756.1/SE6	γ-1	S/BSL	FJ969899.1/GC07.1	γ-12	S/VW (PL)	AY468415.1/FA120	γ-X20	S/BSL
JX429975/GC23	γ-1	S/HS IC	FJ969903.1/GC09	γ-12	S/SeK(PL)	AF121428.1/FA6	γ-X21	S/HS IC
FJ969906.1/GC11	γ-2	S/BSL (PL)	FJ969904/GC08	γ-12	S/AK (PL)	AF252606.1/FA43	γ-X21	S/HS IC
JQ612578/GC21	γ-2	S/AK	FJ969905.1/GC10	γ-12	S/HS IC	DQ418469.1/FA153	γ-X21	S/HS IC
AF217672.1/FA27	γ-3	S/HS IS	FJ969911.1/GC14	γ-12	S/AK	FJ947082.1/GC03	μ-PV	S/HS IC
AY040277.1/FA58	γ-3	S/HS IC	JF906526.1/SE1	γ-12	S/SCC-AK			
AF121432.1/FA11	γ-7	S/HS IC	JN129833.1/SE28	γ-12	S/SCC-AK			

†Epithelium: M, mucosal epithelium; S, skin epithelium. Source: AK, actinic keratosis; ASCUS, atypical squamous cells of undetermined significance; BCC, <definition?>; BD, Bowen's disease; BSL, benign skin lesion (keratoacanthoma, fibroma, benign adnexal tumour); Ca, carcinoma; EV, <definition?>; HS IC, healthy skin from immune competent individual; HS IS, healthy skin from immune suppressed individual; NR, not reported; (PL), perilesional sample; ~~Scar~~, <definition?>; SCC, squamous cell carcinoma; SCC-AK, pooled biopsies of SCC and AK; SeK, seborrheic keratosis; SoK, solar keratosis; VW, viral wart.

‡Recently fully characterized viruses: HPVX13 (β-2), HPV-159 variant; HPVX14b (β-2), HPV-174 variant; RTRX3 (β-2), KC176801 variant; GC15 (γ-1), HPV-163 subtype; FA69 (γ-10), KC108722 variant; SE61 (γ-11), JF966374 variant; FADI3 (γ-11), HPV-154 variant.

CUT region, Fig. 1b, c: β-1=9, β-2=15, γ-7=16, γ-8=8, γ-10=7, γ-11=9, γ-12=24; MY region, Fig. 1b: β-1=9, β-2=17). This may be explained by the poor taxon sampling in these species (Hedtke *et al.*, 2006; Zwickl & Hillis, 2002) and highlights the need to progress in the identification of the missing taxa in order to have a complete picture of the diversity of the genera *Betapapillomavirus* and *Gammapapillomavirus*. However, poor taxon sampling does not seem to be the case for the genus *Alphapapillomavirus*, taking into account the few novel putative HPV strains described until now. In this survey, putative HPV types found in the genus *Alphapapillomavirus* belonged to species containing low-risk mucosal types (α-3 and α-13) or mainly skin associated types (α-2), suggesting that all high-risk mucosal types have been identified. Thus, the phylogenetic incongruences observed in the genus *Alphapapillomavirus* (Fig. S1) could be explained by alternative evolutionary mechanisms such as within-host virus duplication, viral sorting, or viral adaptation after a host switch (Gottschling *et al.*, 2011), in addition to ancestral recombination events (Narechania *et al.*, 2005; Varsani *et al.*, 2006). However, we cannot exclude the possibility that the information content present in the L1 subgenomic fragments may influence the phylogenetic relationships with respect to those obtained with the L1 ORF phylogeny.

Since the phylogenetic analysis of HPV subgenomic amplicons performed by Forslund ~~*et al.*~~ (2007), the number of characterized genomes of HPV from the genera *Betapapillomavirus* (25 HPVs previously known versus 45 viruses currently known) and *Gammapapillomavirus* (nine HPVs previously known versus 54 viruses currently known) has increased considerably (de Villiers, 2013). Although many putative HPV types have been fully characterized so far, there are still a large number of novel genomes to be described. In fact, only putative HPV types obtained with the FAP primers (36 β-PV and 97 γ-PV putative HPV types) were previously analysed (Forslund, 2007). In that sense, this report updates the knowledge of the remaining putative HPV types to be characterized, including those strains identified with other primer systems, and their phylogenetic associations.

This survey demonstrated that the putative HPV types widely outnumber the HPV types that have been completely sequenced, with the existence of at least 202 potential viruses whose genomes need to be fully characterized. Most subgenomic sequences identified belonged to the genus *Gammapapillomavirus*, and many of them were segregated in the currently defined species (de Villiers, 2013). Moreover, we could infer 21 additional putative species inside this genus which are supported by two different phylogenetic inference methods. In conclusion, our data highlight the need to progress in the identification of the missing taxa in order to elucidate the evolution and the medical implications of members of the family *Papillomaviridae*. This knowledge is required to address important questions, such as the definition of putative sequence signatures able to understand the differential tropism observed in some HPV types at the molecular level.

Acknowledgements

We would like to thank Marcela Culasso, María Robson, Mariana de Sanctis and Geraldine Raimundo (Departamento de Inglés, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario) for langauge correction and style. This work was funded by the Agencia Nacional de Promoción Científica y Tecnológica (grants PICT 2008-0421 and PICT 2012-0652). The funders had no role in the study design, data collection and analysis, or in the preparation of the manuscript or the decision to publish. The authors declare that they have no competing interests.

References

Antonsson, A., Forslund, O., Ekberg, H., Sterner, G. & Hansson, B. G. (2000). The ubiquity and impressive genomic diversity of human skin papillomaviruses suggest a commensalic nature of these viruses. *J Virol* **74**, 11636–11641.

Asgari, M. M., Kiviat, N. B., Critchlow, C. W., Stern, J. E., Argenyi, Z. B., Raugi, G. J., Berg, D., Odland, P. B., Hawes, S. E. & de Villiers, E. M. (2008). Detection of human papillomavirus DNA in cutaneous squamous cell carcinoma among immunocompetent individuals. *J Invest Dermatol* **128**, 1409–1417.

Berkhout, R. J., Bouwes Bavinck, J. N. & 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.

Bernard, H. U., Burk, R. D., Chen, Z., van Doorslaer, K., zur Hausen, H. & de Villiers, E. M. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* **401**, 70–79.

Bottalico, D., Chen, Z., Dunne, A., Ostoloza, J., McKinney, S., Sun, C., Schlecht, N. F., Fatahzadeh, M., Herrero, R. & other authors (2011). The oral cavity contains abundant known and novel human papillomaviruses from the *Betapapillomavirus* and *Gammapapillomavirus* genera. *J Infect Dis* **204**, 787–792.

Chen, A. C., McMillan, N. A. & Antonsson, A. (2008). Human papillomavirus type spectrum in normal skin of individuals with or without a history of frequent sun exposure. *J Gen Virol* **89**, 2891–2897.

Chouhy, D., Gorosito, M., Sánchez, A., Serra, E. C., Bergero, A., Fernandez Bussy, R. & Giri, A. A. (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.

Chouhy, D., Bolatti, E. M., Piccirilli, G., Sánchez, A., Fernandez Bussy, R. & Giri, A. A. (2013). Identification of human papillomavirus type 156, the prototype of a new human gammapapillomavirus species, by a generic and highly sensitive PCR strategy for long DNA fragments. *J Gen Virol* **94**, 524–533.

de Villiers, E. M. (2013). Cross-roads in the classification of papillomaviruses. *Virology* [Epub ahead of print]

de Villiers, E. M., Fauquet, C., Broker, T. R., Bernard, H. U. & zur Hausen, H. (2004). Classification of papillomaviruses. *Virology* **324**, 17–27.

Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969–1973.

Forslund, O. (2007). Genetic diversity of cutaneous human papillomaviruses. *J Gen Virol* **88**, 2662–2669.

Forslund, O., Antonsson, A., Nordin, P., Stenquist, B. & Hansson, B. G. (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**, 2437–2443.

Forslund, O., Ly, H. & Higgins, G. (2003). Improved detection of cutaneous human papillomavirus DNA by single tube nested 'hanging droplet' PCR. *J Virol Methods* **110**, 129–136.

Forslund, O., Iftner, T., Andersson, K., Lindelof, B., Hradil, E., Nordin, P., Stenquist, B., Kirnbauer, R., Dillner, J., de Villiers, E. M. & Viraskin Study Group (2007). Cutaneous human papillomaviruses found in sun-exposed skin: Beta-papillomavirus species 2 predominates in squamous cell carcinoma. *J Infect Dis* **196**, 876–883.

Gottschling, M., Göker, M., Stamatakis, A., Bininda-Emonds, O. R., Nindl, I. & Bravo, I. G. (2011). Quantifying the phylodynamic forces driving papillomavirus evolution. *Mol Biol Evol* **28**, 2101–2113.

Harwood, C. A., Surentheran, T., McGregor, J. M., Spink, P. J., Leigh, I. M., Breuer, J. & Proby, C. M. (2000). Human papillomavirus infection and non-melanoma skin cancer in immunosuppressed and immunocompetent individuals. *J Med Virol* **61**, 289–297.

Harwood, C. A., Surentheran, T., Sasieni, P., Proby, C. M., Bordea, C., Leigh, I. M., Wojnarowska, F., Breuer, J. & McGregor, J. M. (2004). Increased risk of skin cancer associated with the presence of epidermodysplasia verruciformis human papillomavirus types in normal skin. *Br J Dermatol* **150**, 949–957.

Hazard, K., Andersson, K., Dillner, J. & Forslund, O. (2007a). Human papillomavirus subtypes are not uncommon. *Virology* **362**, 6–9.

Hazard, K., Karlsson, A., Andersson, K., Ekberg, H., Dillner, J. & Forslund, O. (2007b). Cutaneous human papillomaviruses persist on healthy skin. *J Invest Dermatol* **127**, 116–119.

Hedtke, S. M., Townsend, T. M. & Hillis, D. M. (2006). Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Syst Biol* **55**, 522–529.

Kocjan, B. J., Jelen, M. M., Maver, P. J., Seme, K. & Poljak, M. (2011). Pre-vaccination genomic diversity of human papillomavirus genotype 6 (HPV 6): a comparative analysis of 21 full-length genome sequences. *Infect Genet Evol* **11**, 1805–1810.

Manos, M. M., Ting, Y., Wright, D. K., Lewis, A. J., Broker, T. R. & Wolinski, S. M. (1989). Use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. *Cancer Cells* **7**, 209–214.

Menzo, S., Monachetti, A., Trozzi, C., Ciavattini, A., Carloni, G., Varaldo, P. E. & Clementi, M. (2001). Identification of six putative novel human papillomaviruses (HPV) and characterization of candidate HPV type 87. *J Virol* **75**, 11913–11919.

Narechania, A., Chen, Z., DeSalle, R. & Burk, R. D. (2005). Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses. *J Virol* **79**, 15503–15510.

Peyton, C. L. & Wheeler, C. M. (1994). Identification of five novel human papillomavirus sequences in the New Mexico triethnic population. *J Infect Dis* **170**, 1089–1092.

Pfister, H., Fuchs, P. G., 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.

Shamanin, V., Glover, M., Rausch, C., Proby, C., Leigh, I. M., zur Hausen, H. & de Villiers, E. M. (1994). Specific types of human papillomavirus found in benign proliferations and carcinomas of the skin in immunosuppressed patients. *Cancer Res* **54**, 4610–4613.

Shamanin, V., zur Hausen, H., Lavergne, D., Proby, C. M., Leigh, I. M., Neumann, C., Hamm, H., Goos, M., Haustein, U. F. & other authors (1996). Human papillomavirus infections in nonmelanoma skin cancers from renal transplant recipients and nonimmunosuppressed patients. *J Natl Cancer Inst* **88**, 802–811.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739.

Varsani, A., van der Walt, E., Heath, L., Rybicki, E. P., Williamson, A. L. & Martin, D. P. (2006). Evidence of ancient papillomavirus recombination. *J Gen Virol* **87**, 2527–2531.

Zwickl, D. J. & Hillis, D. M. (2002). Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol* **51**, 588–598.

Dear Authors,

Please find enclosed a proof of your article for checking.

When reading through your proof, please check carefully authors' names, scientific data, data in tables, any mathematics and the accuracy of references. Please do not make any unnecessary changes at this stage. All necessary corrections should be marked on the proof at the place where the correction is to be made; please mark up the correction in the PDF and return it to us (see instructions on marking proofs in Adobe Reader).

Any queries that have arisen during preparation of your paper for publication are listed below and indicated on the proof.

Please provide your answers when returning your proof.

Please return your proof by email (sgmprod@charlesworth-group.com) within 2 days of receipt of this message.

Query no.	Query
1	'FAP' has been defined as 'Forslund-Antonsson primers'. Please confirm this is correct.
2	In the phrase 'Since the phylogenetic analysis of HPV subgenomic amplicons performed by Forslund *et al.* (Forslund, 2007)' the citation has been changed to 'Forslund *et al*. (2007)'. Is this correct or should both Forslund *et al*. (2007) and Forslund (2007) be cited here?
3	For reference 'de Villiers (2013)' please provide volume and page numbers if it is has now been published.
4	Table 1 'EV Scar' – please add definitions of 'EV' and 'Scar' to the table footnote (separate entries please as 'EV' appears alone elsewhere in the table).
5	Table 1: Entry for 'Epithelium/source' for Z95969.1/DL287 has been changed from 'S/SSC' to 'S/SCC'. Please confirm this is correct. Otherwise please add definition of 'SSC' to table footnote.
6	Table 1: 'BCC' - please add definition of 'BCC' to table footnote.
7	'SCC' in Table 1 has been defined in the table footnote as 'squamous cell carcinoma'. Please indicate if this is correct.

Ordering reprints for SGM journals

As a result of declining reprint orders and feedback from many authors who tell us they have no use for reprints, **SGM no longer provides free reprints to corresponding authors**; instead, corresponding authors will receive two emails:

i) An email including a link to download the published PDF of their paper. You can forward this link to co-authors or others, and they can also use it to download the published PDF. The link can be used up to 25 times. This email will be sent out at around the time your article is published online.

ii) An email including a link to the SGM Reprint Service. You can forward this email to your co-authors if you wish, so that they can order their own reprints directly, or to your finance or purchasing department, if orders are placed centrally. This email will be sent out at around the time that your article is finalized for printing.

When you click on the link in this second email, you will be taken to an order page to place your reprint order. Like most online ordering sites, it is necessary to set up an account and provide a delivery address while placing your order, if you do not already have an account. Once an account and delivery address have been set up, these details will be stored by the system for use with future orders. Payments can be made by credit card, PayPal or purchase order.

As reprint orders are despatched by courier, there is a charge for postage and packing.

SUMMARY

- You can create or update your reprint account at any time at http://sgm-reprints.charlesworth.com/
- You will be sent an email when the reprints of this paper are ready for ordering
- **You cannot order reprints of this paper before this email has been sent**, as your paper will not be in the system
- Reprints can be ordered at any time after publication
- You will also receive an email with a link to download the PDF of your published paper

The reprint ordering details will be emailed to the author listed as the corresponding author on the journal's manuscript submission system. If your paper has been published (the final version, not the publish-ahead-of-print version) but you have not received this notification, email reprints@sgm.ac.uk quoting the journal, paper number and publication details.

If you have any questions or comments about the reprint-ordering system or about the link to your published paper, email reprints@sgm.ac.uk