Accepted Manuscript

Title: Human papillomaviruses 53 and 66: clinical aspects and genetic analysis

Authors: Valeria Cento, Noor Rahmatalla, Massimo Ciccozzi, Alessandra Lo Presti, Carlo Federico Perno, Marco Ciotti

PII: S0168-1702(11)00385-6
DOI: doi:10.1016/j.virusres.2011.09.032
Reference: VIRUS 95566

To appear in: *Virus Research*

Received date: 21-6-2011
Revised date: 22-9-2011
Accepted date: 22-9-2011

Please cite this article as: Cento, V., Rahmatalla, N., Ciccozzi, M., Presti, A.L., Perno, C.F., Ciotti, M., Human papillomaviruses 53 and 66: clinical aspects and genetic analysis, *Virus Research* (2010), doi:10.1016/j.virusres.2011.09.032

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Highlights

- The study investigated the genetic variability of HPV 53 and 66
- Novel mutations were uncovered in both HPV types
- HPV 53 L1 and E6 genes clustered in two main clades
- Some of the novel mutations affected the antigenic index of the L1 protein

Accepted Manuscript

ACCEPTED MANUSCRIPT

Human papillomaviruses 53 and 66: clinical aspects and genetic analysis

Valeria Cento[a], Noor Rahmatalla[a], Massimo Ciccozzi[b], Alessandra Lo Presti[b], Carlo Federico Perno[a, c], and Marco Ciotti [c *]

[a]Department of Experimental Medicine and Biochemical Sciences, University Tor Vergata, Rome, Italy. [b]Department of Infectious, Parasitic and Immuno mediated Disease, Istituto Superiore di Sanita', Rome, Italy. [c]Foundation Polyclinic Tor Vergata, Viale Oxford81, 00133 Rome, Italy.

***Correspondence to:** Dr. Marco Ciotti, Laboratory of Molecular Virology, Foundation Polyclinic Tor Vergata, Viale Oxford 81-00133,Rome, Italy.

Tel.: +39 06 20902087; Fax: +39 06 20900413; e-mail: marco.ciotti@ptvonline.it

Accepted Manuscript

ACCEPTED MANUSCRIPT

Abstract

Variants of HPV types may have different oncogenic potential. While HPV 16 and 18 variants have been extensively studied, little is known on the less frequent high-risk types such as HPV 53 and 66. Here, we analyzed the genetic variability of HPV 53 and 66 by sequencing the *E6*, *E7*, *L1* genes and the Long Control Region sequences of HPV 53 and HPV 66 from infected women. Fisher's exact-test was performed to correlated viral variants with cervical lesions. Higher-order interactions among identified mutations were analysed by co-variation and cluster analyses. Antigenic-index alterations following L1 mutations were predicted by Jameson-Wolf algorithm.

In HPV53, novel variants were identified in *L1* (N=9) and *E6* (N=1) genes. The novel L1 mutation P432L was statistically associated with L-SIL lesions (P=0.04) and its development reduced the L1 predicted antigenicity (up to -2.3 for Glu433). HPV 53 *E6* and *L1* sequences clustered phylogenetically into two main clades.

In HPV 66, novel polymorphisms were identified in *L1* (N=4) and *E6* (N=4) genes. The L1 protein mutations S405P and D458N were exclusively found in patients with L-SILs. Seven *E7* variants and 10 LCR variants were for the first time analyzed.

Novel HPV 53 and 66 variants were identified in this study. Some of these mutations were significantly associated with L-SIL lesions and affected the antigenic index of the L1 protein with possible interesting implications in vaccine design.

Key words: HPV 53, HPV 66, genetic analysis

Accepted Manuscript

ACCEPTED MANUSCRIPT

1. Introduction

Oncogenic human papillomaviruses (HPVs) are the recognized etiological agents of cervical cancer. According to the 2010 classification, the family *Papillomaviridae* comprises 120 human types classified in five genera (Alpha-, Beta-, Gamma-, Mu- and Nu-HPVs) (Bernard et al., 2010).

Eighteen types classified as α-papillomaviruses are of particular interest because often found associated with ano-genital cancer. Of these, 15 are considered high-risk types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73, and 82) and 3 were classified as probable high-risk types (26, 53, and 66) (Munoz et al., 2003). Among the high-risk types, HPV 16 and 18 is the most representative and account for about 70% of all cervical cancers (http://www.who.int./hpvcentre). Each of HPV 16 and 18 forms are taxonomic clustered together with six other HPV types in two different branches called the α-9 and α-7 species (de Villiers et al., 2004). A third species (α-6) includes the HPV types 66, 53, 56 and 30.

An HPV type is considered as such when the nucleotide sequence of its *L1* gene differs from that of any other type by at least of 10%. The term subtype is instead used to define those HPVs whose *L1* sequence differs from that of the closest one of 2-10%. Finally, variants of HPV types differ by less than 2% of their *L1* nucleotide sequence (de Villiers et al., 2004).

The prevalence of HPV variants may differ in different parts of the world (Yamada et al., 1997; Cento et al., 2009; Calleja-Macias et al., 2005). Furthermore, geographically defined HPV 16 and 18 variants are associated with a higher risk of persistence and oncogenicity (Sichero et al., 2007). While the characteristics of α-9 and α-7 species, to which HPV 16 and HPV 18 belong, have been extensively studied, little is known about the genetic features of the high risk HPV types 53 and 66, belonging to the α-6 species. The aim of this study was to describe the genetic variability of HPV 53 and 66. All viral sequences obtained from patients included in the study were analyzed to identify and characterize novel polymorphisms and new viral variants. A phylogenetic approach was used to define the presence and localization of viral genomic regions and/or codons subject to positive selective pressure, as well as to identify significant clustering of HPV 53 and 66 variants.

ACCEPTED MANUSCRIPT

The Long Control Region (LCR) of both HPV types was also examined, looking for nucleotide variations that may affect potential binding sites.

The knowledge of mutant viral proteins may improve the understanding of molecular mechanisms underlying disease progression and the vaccination approach.

Accepted Manuscript

ACCEPTED MANUSCRIPT

2. Material and methods

2.1. Samples collection

Exo- and endocervical cells were collected from women attending the Gynecology and Colposcopy Unit of Tor Vergata University Hospital from January 2007 to August 2010. The collected cells were placed in PreservCyt solution (Hologic, Marlborough, MA) and stored, according to the manufacturer's instructions, until analysis. The study included women mono-infected with HPV 53 or 66 with either normal cytology (11 women) or abnormal pap smears: 23 with atypical cells of undetermined significance (ASCUS), and 20 with low-grade squamous intraepithelial lesions (L-SIL). The demographic and clinical features of the women examined as well as the number of *E6*, *E7* and *L1* genes and LCR fragments analyzed are reported in Table 1. Informed consent was obtained from all women enrolled in the study.

2.2. Amplification and sequencing of L1, E6, E7 genes and LCR of HPV 53 and 66

The *L1* gene of HPV 53 and 66 was amplified using the degenerate primers MY09/MY11 (Manos et al., 1989). For *E6*, *E7* and LCR, individual amplification protocols were developed and are reported in Table S1. Five µl of DNA were added to 45µl of reaction mix containing 1X PCR buffer, 200 µM dNTPs, 2 mM $MgCl_2$, 40 pmol of each primer, and 1U of Taq Gold polymerase (Applied Biosystems, Monza, Italy). PCR products were run on a 2% agarose gel. For all PCR positive samples, sequencing of *L1*, *E6*, *E7* and LCR was performed with the Big Dye Terminator V.3.1 cycle sequencing kit (Applied Biosystem, Monza, Italy) and with the same reverse and forward primers used in the amplification reaction. The obtained sequences were subtyped and numbered according to the reference sequences of HPV 53 and HPV 66 genotypes (GenBank accession numbers: X74482 for HPV 53 and U31794 for HPV 66). All HPV 53 and 66 sequences derived from this study have been deposited in GenBank under the following accession numbers: JN661387- JN661412 (HPV 53 E6); JN661414- JN661436 (HPV 66 E6); JN661437- JN661462

Accepted Manuscript

ACCEPTED MANUSCRIPT

(HPV 53 E7); JN661463- JN661483 (HPV 66 E7); JN661484- JN661505 (HPV 53 L1); JN661506-JN661527 (HPV 66 L1); JN661528- JN661551 (HPV 53 LCR); JN661552- JN661558 (HPV 66 LCR).

Multiple sequence alignments of the full length *E6, E7* and LCR regions as well as of the 450 bp fragment of the *L1* gene were performed. The reference sequences added to each alignment are reported in Table S2. Alignments were performed using Clustal W Multiple Sequences Alignment software, and manually edited with the BioEdit software (Hall, 1999). In particular, all coding sequences were firstly aligned at the amino acid level, in order to account for possible in-frame insertions and/or deletions. Afterwards, amino acid sequences were re-converted in the original nucleotides. All subsequent analyses were then performed at nucleotide level.

2.3. Genetic variability analysis

Evolutionary divergence of HPV 53 and HPV 66 sequences was estimated as the extent of nucleotide substitutions per site determined by the HKY model of MEGA v5.0 (Tamura et al., 2011) applying a gamma distribution with shape parameter = 0.8. CodonFreq F3x4 was chosen and gaps were treated as "missing data" (cleandata=0).

The dN/dS rate was estimated by Maximum Likelihood (ML) approach implemented in the program HyPhy (Pond et al., 2005). In particular, the global (assuming a single selective pressure for all branches) and site specific positive selection was estimated by random effect likelihood (REL) model, a variant of the Nielsen-Yang approach (Nielsen and Yang, 1998), which assumes that exists a discrete distribution of rates across sites and allows both dS and dN to vary site-by-site independently (Kosakovsky et al., 2005). In order to select sites under selective pressure we assumed a p value of ≤ 0.05 or a posterior probability of ≥ 0.9 (Kosakovsky et al., 2005). The newik user tree was generated in MEGA5 (HKY model, shape parameter = 0.8, CodonFreq F3x4, cleandata=0).

ACCEPTED MANUSCRIPT

Identical sequences were removed before calculating genetic variability parameters.

2.4 Phylogenetic analysis

The Bayesian phylogenetic tree was reconstructed by means of MrBayes, using the HKY 85 model of nucleotide substitution, a proportion of invariant sites, and gamma distributed rates among sites. A Markov Chain Monte Carlo (MCMC) search was made for 10 X 10^6 generations using tree sampling every 100th generation and a burn-in fraction of 50%. Statistical support for specific clades was obtained by calculating the posterior probability of each monophyletic clade, and a posterior consensus tree was generated after a 50% burn-in. Clades with a posterior probability of 1 were considered significant.

The coefficients of differentiation among clades identified by phylogenetic analyses of HPV 53 and HPV 66 L1 and E6 gene were calculated using MEGA version 5.0 software (Tamura et al., 2011), under the same nucleotide/amino acid substitution models used for phylogenetic trees inference. Identical sequences were removed before calculating phylogenetic relationships.

2.5 Transcription factors binding site analysis

The MATCH^{TM} software (Kel et al., 2003) was used to search for potential binding sites for cellular and viral transcriptional factors within the LCR of the two HPV types analyzed: AP-1, E2, GRE, NF-1, Oct-1, TATA, YY1, C/EBP, Sp1, SRY, AML-1a and c-Myc/c-Max. Cut-off values and coincidence levels between consensus binding sites and the LCR sequence of each type were adjusted in order to minimize both the number of negative and positive faults.

2.6 Antigenic index calculation

Translated L1 sequences containing selected mutations were analyzed with the Jameson-Wolf algorithm in the Lasergene Protean v8.1 program (DNASTAR Inc., Madison, WI) to predict the antigenic index (Ai). This algorithm integrates several parameters to calculate the antigenicity of the

Accepted Manuscript

ACCEPTED MANUSCRIPT

sequence based on the characteristics of its primary amino acid chain: hydrophilicity (Hopp-Woods), surface probability (Emini), flexibility of the protein backbone (Karplus-Schulz), and secondary structure prediction (Chou-Fasman and Garnier), with regions of positive Ai value indicating possible antigenic determinants (Jameson and Wolf, 1988).

2.7 Statistical analysis

Fisher's exact test was performed to statistically assess the association of *L1*, *E6* and *E7* mutations with different grades of cytological cervical lesions. All already described mutations and the novel mutations identified with prevalence >20% or in >1 patients were included in the analysis.

For each pair of mutations and corresponding wild-type residues, Fisher's exact test was also performed to assess whether co-occurrence of the mutated residues differed significantly from what would be expected under an independence assumption. The Benjamini-Hochberg method was used to correct for multiple testing, at a false discovery rate of 0.05. Samples having a mixture of two or more mutations at a given pair of positions were ignored when calculating the co-variation, due to the impossibility of verifying whether these mutations are indeed located in the same viral genome.

In order to identify and summarize higher-order interactions of mutations, we transformed the pairwise phi correlation coefficients into dissimilarity values. Based on these pairwise dissimilarity values, a dendrogram was computed by hierarchical clustering. Finally, the stability of the resulting dendrogram was assessed from 100 bootstrap replicates. The details of this explorative data analysis procedure have been described elsewhere (Svicher et al., 2009). All statistical analyses were carried out with R software.

Accepted Manuscript

ACCEPTED MANUSCRIPT

3. Results

3.1 Study population

Of the study population examined, 30 women were infected by HPV 53, and 24 by HPV 66 (Table 1). No co-infections by other HPV types were detected. In our study population neither of these two HPV types was found associated with high-grade squamous intraepithelial lesions (H-SIL) or cervical cancer. In particular, cytological examination of cervical smears was negative for the presence of abnormal cells in 9/30 (30.0%) women infected with HPV 53 and in 2/24 (8.3%) women infected with HPV 66. ASCUS lesions were slightly more frequent in patients infected with HPV 53 (15/30, 50.0%) than in women positive for HPV 66 (8/24, 33.3%), but this difference was not statistically significant. On the contrary, L-SIL lesions were significantly more frequently detected in women infected with HPV 66 (14/24, 58.3%) than in those with HPV 53 (6/30, 20.0%) (P=0.005).

3.2 HPV 53 genomic diversity and phylogenetic analysis

The viral sequences of *L1*, *E6*, *E7* genes and LCR of HPV 53 were analyzed in terms of nucleotide and amino acidic variability.

The *E6* gene showed a higher mean number of nucleotide substitutions per site [Mean (Standard Error, SE)=0.607 (0.392-0.889)] than those of *L1* [Mean (SE)=0.021 (0.004)], *E7* [Mean (SE)=0.004 (0.002)] or *LCR* [Mean (SE)=0.016 (0.004)], but this difference did not reach statistical significance (data not shown).

Overall, dN/dS values were low for all genomic regions analyzed, showing no evidences of a strong positive selection pressure acting on the virus. The median dN/dS ratio was slightly higher for the *E6* gene [Mean (95% Confidence interval, CI) dN/dS=0.607 (0.392-0.889)], respect to either *L1* [Mean (95% CI) dN/dS=0.383 (0.279-0.510)] and *E7* [Mean (95% CI) dN/dS=0.348 (0.058-1.077)] regions. No sites under positive selection were identified neither in both *E6* and *E7* genes. On the

ACCEPTED MANUSCRIPT

contrary, positive selection was found at the level of L1 protein. In particular, positions 360 and 432 showed a dN/dS ratio >1, though with low absolute values (range 1.39 and 1.45, respectively), indicating a low strength of diversifying selection acting on the protein.

The phylogenetic trees calculated from either *E6* or *L1* sequences showed two main clades (Prado et al., 2005), Fig. 1, panels A and B. No geographical associations were observed among the viral variants and the Italian sequences were interspersed among clades. The analysis was repeated and confirmed also for *E7* sequences (data not shown).

The coefficients of differentiation among clades were 0.470 (Standard Error, SE=0.091) substitutions/site for *E6* gene and 0.557 (SE=0.138) substitutions/site for *L1* gene.

No association between the main clades and the grade of cervical lesion was found. Indeed, the ASCUS lesions were present in 8/15 (53.3%) patients belonging to clade A and in 7/15 (47.7%) patients from clade B, while L-SILs were present in 4/15 (26.7%) and 2/15 (13.3%) patients from clades A and B, respectively.

3.3 HPV 53 L1 sequence variations and prediction of antigenic index

Nucleotide and amino acids analysis of the 450 bp fragment of the *L1* gene from 22 patients infected with HPV 53 revealed the presence of 17 different nucleotide mutations, with 8 (47.1%) of these leading to an amino acid change in the L1 protein, Table 2, panel A. Nine out of 17 nucleotide mutations identified have never been described before. Among these novel mutations, 7 led to the four amino acid changes S343C, Y360L, P432L and N447D. The P430S mutation has been described elsewhere (Prado et al., 2005; Wyant et al., 2011). This mutation was exclusively detected in sequences belonging to clade A (P= 0.001 with Fisher exact test after Benjamini-Hockberg correction). By co-variation analysis, the P430S was found to be positively associated with the synonymous nucleotide mutations at position 6677 (ACT to ACA), 6704 (TCT to TCC), 6740 (AGA to AGG), 6908 (GTA to GTG), and 6998 (TTG to TTA) (all phi values comprised between 0.6 and 1; P-values for associations always <0.004, data not shown). All sequences

Accepted Manuscript

ACCEPTED MANUSCRIPT

showing this variations' pattern belonged to clade A. This association was also confirmed by performing an average linkage hierarchical agglomerative cluster analysis, able to investigate the co-evolution of mutations in more detail, Fig. 3, panel A.

By comparing amino acid sequences of the L1 protein of HPV 16 and HPV 53, it was possible to localize single amino acid mutations of HPV 53 L1 on the secondary structure of the protein, Fig. 2. The novel amino acid mutations S343C, Y360L and N447D were located in the β-H2, β-I and β-J strands, respectively. On the contrary, the P430S and the P432L were located in the H4 - β-J connecting region (Chen et al., 2000; Bishop et al., 2007).

Among the novel amino acid mutations identified, the Y360L, N447D and P432L were exclusively found in patients with L-SIL, Table 2, panel A. This association was statistically significant for P432L (P=0.04, Fisher exact test).

By predicting the antigenic index of mutant sequences, the P432L was also associated with a wide antigenic alteration of the surrounding L1 amino acids from Ile 423 to Pro 437. In particular, surface probabilities were strongly affected in the region from Gln 429 to Lys 434 with a range of reduction from -1.99 up to -2.56 for the downstream Glu 433 and Lys 434, while the antigenic indexes showed a range of negative alterations from Gln 429 to Pro 437 from -0.34 up to -2.3 for Glu 433, Table S3. Similarly Y360L mutations was found to increase the calculated hydrophobicity of L1 residues 356 to 364 (range of increase: 0.56-0.57) and to reduce the surface probability of residues 357-362 (rage of reduction: -0.59 up to -1-43 for Ile 357), Table S3. As a consequence, a wide antigenic alteration was observed following Y360L mutation, which caused a decrease in the antigenic index values of amino acids from Thr 351 to Arg 362 (range of decrease: from -0.35 up to -1.8 for Lys 355). On the contrary, S343C, P430S and N447D substitutions led to minor modifications in hydrophobicity and surface probability, with poor or null consequences on antigenic index of the L1 protein (data not shown), although oriented toward negative alterations - i.e., acquisition of more hydrophobic characteristics.

Accepted Manuscript

ACCEPTED MANUSCRIPT

3.4 HPV 53 E7 and E6 sequence variations

Analysis of the full-length *E7* gene in 26 patients infected with HPV 53 revealed the presence of 3 nucleotide substitutions, Table 2, panel B. Of these, two led to amino acid changes at positions 79 (S79P, 19.2% prevalence) and 85 (I85M, 30.8% prevalence). All detected mutations were exclusively present in sequences belonging to clade A. Overall, the *E7* sequences from 17/26 patients (65.4%) showed no variations with respect to the wild-type reference strain.

All 27 patients analyzed contained at least one nucleotide mutation in *E6* gene, and a total of 19 variations were identified, with 10 of them leading to amino acid changes in the E6 protein, Table 2, panel C. Only one novel nucleotide substitution was detected, which led to the amino acid change A149S (2/27 patients, 7.4% prevalence). The majority of amino acid changes (Q4R, E10Q, K25Q, T25K and Y82H) were representative of clade A, as they were never detected in any sequence belonging to clade B ($P<0.001$ for all mutations with Fisher exact test after Benjamini-Hockberg correction).

Co-variation analysis performed on *E6* sequences highlighted the presence of very strong positive association among specific mutations (all phi values comprised between 0.7 and 1; P-values for associations always <0.001, data not shown), which resulted to correspond to two specific co-variation patterns as identified by the average linkage hierarchical agglomerative cluster analysis, Fig. 3, panel B. Mutations clustered according to phylogenetic classification. Indeed, the topology of the dendrogram highlighted a very strong correlation among all clade A mutations, that resulted clearly separated from clade B mutations.

No specific association was observed between the *E7* or *E6* sequence variations and the cytological lesion.

3.5 HPV 53 Long Control Region sequence variations

Accepted Manuscript

ACCEPTED MANUSCRIPT

A total of 17 nucleotide substitutions were observed in the region spanning from nucleotides 7422 to 7810, Table 2, panel D. The change found most frequently was the C to T transition at position 7428 (15/30 patient, 50.0% prevalence). No specific association was observed between nucleotide change and grade of cytological lesions.

MATCH™ analysis revealed along the region analyzed a series of potential binding sites for the following transcription factors: SBF-1, Pax-6, c-Myb, AGL3, E2F, E2, FOXD3, AP-1 and Elf-1. Several mutations were located in correspondence of these sites. It remains to be established whether such mutations affect the function of these potential binding sites.

3.6 HPV 66 genomic diversity and phylogenetic analysis

The LCR showed the higher mean number of nucleotide substitutions per site [mean (SE)=0.041 (0.006)] than those of HPV 66 *L1* [mean (SE)= 0.011 (0.003)], *E6* [mean (SE)= 0.009 (0.003)] and *E7* [mean (SE)= 0.018 (0.006)], but this difference did not reach statistical significance (data not shown), probably due to the low number of LCR sequences available. Overall, dN/dS values were higher for *E6* [Mean (95% CI) dN/dS =0.618 (0.356-0.985)] and *E7* sequences [Mean (95% CI) dN/dS=1.277 (0.613-2.270)] than for *L1* sequences [Mean (95% CI) dN/dS=0.291 (0.146-0.511)]. Nevertheless, no sites under positive selective pressure were identified.

The phylogenetic tree for *E6* and *L1* genes of HPV 66 was inferred using the *E6* and *L1* sequences obtained from this study along with those retrieved from GenBank (Fig. 4).

Also in this case, two main clade of phylogenetically related strains were identified on either *L1* and *E6* sequences (Fig. 4, panels A and B), but no specific association with variants from different geographical areas or cytological cervical lesion was observed. Indeed, ASCUS lesions were present in 4/7 (57.1%) patients belonging to clade A and in 4/17 (23.5%) patients from clade B, while L-SILs were present in 2/7 (28.6%) and 12/17 (70.6%) patients from clades A and B, respectively.

ACCEPTED MANUSCRIPT

The coefficients of differentiation among clades were 0.692 (Standard Error, SE=0.118) substitutions/site for L1 gene and 0.278 (SE=0.000) substitutions/site for E6 gene.

3.7 HPV 66 L1 sequence variations

Nucleotide and amino acids analysis of the 450 bp fragment of the *L1* gene from 22 patients infected with HPV 66 revealed the presence of 11 nucleotide variations (Table 3, panel A). The nucleotide changes at positions 6855 (GGC to GGT), 6859 (TCC to CCC), and 7016 (CTG to CTA) were identified for the first time in the present study. In addition, a novel three-nucleotide insertion, leading to a hystidine insertion at position 436 of L1 protein, was detected in one patient with no evidences of cytological lesions. Overall, at least one nucleotide variation was observed in 20/22 (90.9%) patients analyzed, and 16/20 (80.0%) presented more than one mutation.

A positive correlation was identified, by co-variation analysis, among the nucleotide mutations at positions 6927 (CAG to CAA), 6660 (AGC to AGT), 6858 (TTA to TTG), 6849 (AAT to AAC) and 6711 (GCC to GCA) (all phi values comprised between 0.8 and 1; P-values for associations always <0.001, data not shown). All these mutations were exclusively found in sequences belonging to clade A, identified by phylogenetic analysis. The higher-order interaction among these mutations is represented in the dendrogram in Fig. S1.

At the amino acid level, two mutations (S405P and D458N) were identified in 2/22 patients each (9.1% prevalence). Of these, the S405P was a novel mutation and resulted located in the connecting domain between H3 helix and H4 helix in L1 protein secondary structure (Fig. 2). The D458N mutation was instead located in the β-J – H5 connecting domain (Fig. 2). All 4 patients harboring amino acid mutations presented cytological abnormalities diagnosed as L-SILs, although the association with this type of lesion was not statistically significant.

The S405P mutation had little or no effect on the estimated hydrophobicity and antigenic index of the L1 protein (Table S4). Indeed the hydrophobicity was slightly decreased for amino acids 401 to

ACCEPTED MANUSCRIPT

409, while the antigenic indexes showed a 0.4 increase at position 402, a 0.2 decrease at position 403, and a 0.15 increase at position 404. On the contrary, at position 458, the substitution of a negatively charged residue (Asp) with a neutral one (Asn) caused a marked decrease in the calculated antigenic index for amino acids from 455 to 461 (mean change = -0.93) (Table S4). At the same positions also the relative surface probability was decreased, while hydrophobicity was unaffected.

3.8 HPV 66 E7 and E6 sequence variations

To our knowledge, this is the first study analyzing the *E7* genetic variability of HPV 66. Analysis of the full-length *E7* gene in 21 patients infected with HPV 66 revealed the presence of 7 different nucleotide substitutions, Table 3, panel B. Of these, 3 led to amino acid changes in the E7 protein (V5A, 81.0% prevalence; V11I, 81.0% prevalence; S104T, 76.4% prevalence). Overall 17/21 (80.9%) patients showed at least one nucleotide substitution. No statistically significant correlation patterns were found among the *E7* mutations identified.

The full-length *E6* gene was analyzed for 23 patients. A total of 9 different nucleotide substitutions were detected in 20/23 (86.9%) patients analyzed, Table 3, panel C. Among the 9 nucleotide mutations identified, 7 had no impact on the E6 protein, while the mutations at nucleotide positions 360 (TAT to GAT) and 378 (AGT to GGT) led to the amino acid changes Y87D and S93G, respectively. The S93G mutation has been described before (Wyant et al., 2011), while the Y87D is a novel mutation and was detected in 2/23 patient (8.7% prevalence) with ASCUS and L-SIL, respectively.

By co-variation analysis, the synonymous nucleotide mutations at positions 264 (CTA to TTA) and 378 (AGT to GGT) resulted to be positively associated (phy = 0.85, P-value for association = 0.04).

No specific association among *E7* or *E6* sequence variations and cytological lesion was observed.

ACCEPTED MANUSCRIPT

3.9 HPV 66 Long Control Region sequence variations

To our knowledge, the LCR of HPV 66 has not been analyzed in terms of nucleotide variability before. In our study, LCR sequences were obtained for 7 patients and 10 nucleotide mutations were detected, Table 3, panel D. Among these, the most prevalent were those at positions 7620 (T to A, 4/7 patients, 57.1% prevalence), 7438 (C to T, 3/7 patients, 42.9% prevalence), and 7779 (A to T, 3/7 patients, 42.9% prevalence).

MATCH^{TM} analysis revealed the presence of three potential binding sites for the transcription factors Pax-6, E2F and Elf-1. The A to T transversion at position 7779 fell within the Elf-1 binding site. However, when the Elf-1 sequence with this mutation inserted was re-submitted to the MATCH^{TM} program, it did not appear to affect the main recognition sequence for Elf-1 transcription factor.

Accepted Manuscript

ACCEPTED MANUSCRIPT

4. Discussion

This is the first report analyzing extensively the genetic variability of LCR, *L1*, *E6* and *E7* genes of HPV 53 and 66, detected as mono-infection in cervical swabs of Italian women examined at the Gynecological and Colposcopy Unit of Tor Vergata University Hospital.

Both HPV 53 and 66 have been associated with different grades of cervical lesions or cancer (Munoz et al., 2003; Meyer et al., 2001; De Vuyst et al., 2003); however, in this study, none of these two viral types were found in cytological cervical lesions worse than L-SILs. The prevalence of L-SIL lesions was significantly higher for patients infected with HPV 66 than HPV 53 (P =0.005). Overall, the genetic diversity of HPV 53 or HPV 66 was not dependent upon the degree of cervical lesions found for any of the genomic regions analyzed, neither in terms of nucleotide or amino acid variability. Nucleotide and amino acid variability was particularly low in HPV 66 sequences, while specific mutations found in HPV 53 determined a dichotomic branching of *L1* and *E6* sequences. Indeed, based on the phylogenetic analysis of *L1* and *E6* genes of HPV 53, the HPV variants identified in this study fell into two main clade defined as A and B. Dichotomy was already observed analyzing the LCR of HPV 53, 44 and 68 (Prado et al., 2005; Calleja-Macias et al., 2005). Processes such as genetic drift and some positive selection would explain this pattern (Prado et al., 2005). No specific association was found between the specific phylogenetic clade and the grade of cervical lesion. However, more data on a larger population are needed to assess the real clinical impact of these results.

In the 450 bp fragment of the HPV 53 *L1* gene, a total of 9 novel nucleotide variations and 4 novel amino acids changes were identified.

Notably, the novel mutations Y360L, P432L and N447D were exclusively present in patients with abnormal cytology and within sequences belonging to clade A. The P432L, localized in the H4-BJ interconnection domain of L1 protein, was statistically associated with L-SIL lesions and was also found to increase the hydrophobicity and reduce the antigenicity of the surrounding L1 protein region. This result may provide, with the easier escape from host's immune system control, an

ACCEPTED MANUSCRIPT

explanation for the relatively higher pathogenicity of this variant. To a lesser extent, also the known P430S mutation was found to alter the antigenicity of the surrounding L1 region. This position was also found to be under a positive selective pressure, suggesting a possible fixation of the mutation in the general population.

In the HPV 53 E6 protein, only 1 novel nucleotide variation was detected, leading to the non-conservative A149S substitution. However, the lack of information on the structure of HPV 53 E6 protein and the diversity among the E6 proteins of HPV 53 and HPV 16, did not allow envisaging a potential effect of this and other mutations identified on the function of E6 protein.

The HPV 66 analysis brought to the uncovering of 3 novel nucleotide mutations in the *L1* gene. Of these, one resulted in the S405P amino acid change. In addition, a novel 3-bases insertion, determining the insertion of a hystidine at position 436 of L1 protein, was also detected. The two amino acid mutations identified in the L1 protein were located in connecting domains and were exclusively detected in patients with L-SIL lesions. The D458N mutation was found to cause a marked decrease in the calculated antigenic index, thus suggesting a different pathogenicity of this mutant protein.

In the *E6* gene, 4 novel nucleotide variations were detected with one of them resulting in an amino acid change in the E6 protein. Nucleotide variations of the *E7* gene and LCR of HPV 66 were for the first time described in this study. In the *E7* gene, a total of 7 different nucleotide substitutions were detected, with 3 affecting the E7 protein. As expected, the LCR was found to be more variable, with 10 nucleotide variations detected. Substitutions within the LCR may influence the location and binding affinity of cellular and viral transcriptional factors binding sites. Indeed, one of the detected mutations was localized in the Elf-1 binding site. However, the effect of this mutation on the transcriptional factors binding sites was not investigated in this study.

Considering all mutations detected in the *E7*, *E6* and LCR sequences of HPV 66, no specific associations were found between them and the cervical dysplasia, probably due to the relatively low number of patients studied.

ACCEPTED MANUSCRIPT

A specific correlation between HPV 16 and 18 variants and higher risk of persistence and oncogenicity has been described (Sichero et al., 2007), whereas no similar data are currently available for less prevalent HPV types. Even considering the limited number of specimens available, specific mutations in the *L1* gene of both HPV 53 and 66 were associated with L-SIL lesions in the present study, though they cannot be considered at the moment predictive of a further worsening of the cervical pathology. None of the HPV 53 and HPV 66 mutations identified have been previously associated with different pathogenetic potential in other high-risk types, neither corresponded to known variants of HPV 16.

Nevertheless, some HPV 53 and 66 L1 protein mutations might alter viral antigenicity, with possible implications in vaccine design (Pastrana et al., 2001). Therefore, larger studies on a population exhibiting a wider spectrum of cervical lesions are necessary to validate this hypothesis.

Acknowledgements

This work was supported by grants from the Italian National Institute of Health, the Italian Ministry of Instruction University & Research (MIUR), "Progetto FILAS" and AVIRALIA Foundation.

Accepted Manuscript

ACCEPTED MANUSCRIPT

References

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 oftaxonomic amendments. Virology 401,70-79.

Bishop, B., Dasgupta, J., Klein, M., Garcea, R.L., Christensen, N.D., Zhao, R., Chen, X.S., 2007. Crystal structures of four types of human papillomavirus L1 capsid proteins: understanding the specificity of neutralizing monoclonal antibodies. J Biol Chem 282, 31803-31811.

Calleja-Macias, I.E., Kalantari, M., Allan, B., Williamson, A.L., Chung, L.P., Collins, R.J., Zuna, R.E., Dunn, S.T., Ortiz-Lopez, R., Barrera-Saldaña, H.A., Cubie, H.A., Cuschieri, K., Villa, L.L., Bernard, H.U., 2005. Papillomavirus subtypes are natural and old taxa: phylogeny of the Human papillomavirus (HPV) types 44/55 and 68a/b. J Virol 79, 6565-6569.

Calleja-Macias, I.E., Villa, L.L., Prado, J.C., Kalantari, M., Allan, B.,Williamson, A.L., Chung, L.P., Collins, R.J., Zuna, R.E., Dunn, S.T., Chu, T.Y., Cubie, H.A., Cuschieri, K., von Knebel-Doeberitz, M., Martins, C.R., Sanchez, G.I., Bosch, F.X., Munoz, N., Bernard, H.U., 2005. Worldwide genomic diversity of the high-risk human papillomavirus types 31, 35, 52, and 58, four close relatives of human papillomavirus type 16. J Virol 79, 13630-13640.

Cento, V., Ciccozzi, M., Ronga, L., Perno, C.F., Ciotti, M., 2009. Genetic diversity of human papillomavirus type 16 E6, E7, and L1 genes in Italian women with different grades of cervical lesions. J Med Virol 81, 1627-1634.

Chen, X.S., Garcea, R.L., Goldberg, I., Casini, G., Harrison, S.C., 2000. Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16. Mol Cell 5, 557-567.

ACCEPTED MANUSCRIPT

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

De Vuyst, H., Steyaert, S., Van Renterghem, L., Claeys, P., Muchiri, L., Sitati, S., Vansteelandt, S., Quint, W., Kleter, B., Van Marck, E., Temmerman, M., 2003. Distribution of human papillomavirus in a family planning population in Nairobi, Kenya. Sex Transm Dis 30, 137-142.

Hall, T.A., 1999. Bioedit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98 NT. Nucleic Acids Symp Ser 41, 95-98.

http://www.who.int/hpvcentre

Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754-755.

Jameson, B.A., Wolf, H., 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. Cabios 4, 181-186.

Kel-Margoulis, O.V., Tchekmenev, D., Kel, A.E., Goessling, E., Hornischer, K., Lewicki-Potapov, B., Wingender, E., 2003. Composition-sensitive analysis of the human genome for regulatory signals. In SilicoBiol 3, 145-171.

Kosakovsky Pond, S.L., Frost, S.D., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22, 1208-1222.

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

Meyer, T., Arndt, R., Beckmann, E.R., Padberg, B., Christophers, E., Stockfleth, E., 2001. Distribution of HPV 53, HPV 73 and CP8304 in genital epithelial lesions with different grades of dysplasia. Int J Gynecol Cancer 11, 198-204.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Muñoz, N., Bosch, F.X., de Sanjosé, S., Herrero, R., Castellsagué, X., Shah, K.V., Snijders, P.J., Meijer, C.J., 2003.Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 348, 518-527.

Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.

Pastrana, D.V., Vass, W.C., Lowy, D.R., Schiller, J.T., 2001. NHPV16 VLP vaccine induces human antibodies that neutralize divergent variants of HPV16. Virology 279, 361-369.

Prado, J.C., Calleja-Macias, I.E., Bernard, H.U., Kalantari, M., Macay, S.A., Allan, B., Williamson, A.L., Chung, L.P., Collins, R.J., Zuna, R.E., Dunn, S.T., Ortiz-Lopez, R., Barrera-Saldaña, H.A., Cubie, H.A., Cuschieri, K., von Knebel-Doeberitz, M., Sanchez, G.I., Bosch, F.X., Villa, L.L., 2005. Worldwide genomic diversity of the human papillomaviruses-53, 56, and 66, a group of high-risk HPVs unrelated to HPV-16 and HPV-18. Virology 340, 95-104.

Pond, S.L., Frost, S.D., Muse S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21, 676-679.

Lazcano-Ponce E, Herrero R, Muñoz N, Cruz A, Shah KV, Alonso P, Hernández P,Salmerón, J., Hernández, M., 2001. Epidemiology of HPV infection among Mexican women with normal cervical cytology. Int J Cancer 91, 412-420.

Sichero, L., Ferreira, S., Trottier, H., Duarte-Franco, E., Ferenczy, A., Franco, E.L., Villa L.L., 2007. High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. Int J Cancer 120, 1763-1768.

Svicher, V., Gori, C., Trignetti, M., Visca, M., Micheli, V., Bernassola, M., Salpini, R., Gubertini, G., Longo, R., Niero, F., Ceccherini-Silberstein, F., De Sanctis, G.M., Spanò, A.,

ACCEPTED MANUSCRIPT

Cappiello, G., Perno, C.F., 2009. The profile of mutational clusters associated with lamivudine resistance can be constrained by HBV genotypes. J Hepatol 50, 461-470.

Swofford, D., Sullivan, J., 2003. Phylogeny inference based on parsimony and other methods with PAUP*. In: Salemi M, Vandamme AM. (Eds.), The phylogenetic handbook—a practical approach to DNA and protein phylogeny. Cambridge University Press, New York, pp. 160–206.

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 [Epub ahead of print]

Wyant PS, Cerqueira DM, Moraes DS, Leite JP, Martins CR, de Macedo Brígido M, Raiol, T., 2011. Phylogeny and polymorphism in the long control region, E6, and L1 of human papillomavirus types 53, 56, and 66 in central Brazil. Int J Gynecol Cancer 21, 222.

Yamada, T., Manos, M.M., Peto, J., Greer, C.E., Munoz, N., Bosch, F.X., Wheeler, C.M., 1997. Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. J Virol 71, 2463.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Figure legends:

Fig. 1 Bayesian phylogenetic trees of HPV 53 L1 (A) and E6 (B) genes. The trees were built with Mr Bayes 3.1 software. Branch lengths were estimated with the best fitting nucleotide substitution model (HKY 85) according to a hierarchical likelihood ratio test and were drawn to scale with the bar at the bottom indicating 0.0080 (A) and 0.01 (B) nucleotide substitutions per site.

The trees were rooted using the midpoint rooting method. One * along the branches represent significant statistical support for the clade subtending that branch (posterior probability>90%).

Fig. 2 Amino acid sequence alignment of HPV 16, HPV 53, and HPV 66 L1 protein. The secondary structural elements identified from the HPV 16 crystal structures are reported above the sequences. HPV 16 sequences reported are from clone 114/K (EU118173). The accession numbers of HPV 53 and HPV 66 are reported. Numbers refer to each HPV sequence. HPV 53 and HPV 66 sequence variations are aligned to the reference sequence of the corresponding genotype.

Secondary structural elements are indicated as reported in (Chen et al., 2000). Identical amino acids are indicated with dots "·". Gaps in the alignment are indicated with dashes "-".

Fig. 3 Dendrogram obtained from average linkage hierarchical agglomerative clustering of HPV 53 L1 **(A)**, and HPV 53 E6 **(B)** mutations. The length of branches reflects distances between genotypes in the original distance matrix. Bootstrap values, indicating the significance of clusters, are reported in the boxes.

Fig. 4 Bayesian phylogenetic trees of HPV 66 *L1* **(A)** and *E6* **(B)** genes. The trees were built with Mr Bayes 3.1 software. Branch lengths were estimated with the best fitting nucleotide substitution model (HKY 85) according to a hierarchical likelihood ratio test and were drawn to scale with the bar at the bottom indicating 0.0040 (A) and 0.0020 (B) nucleotide substitutions per site. The trees were rooted using the midpoint rooting method. One * along the branches represent significant statistical support for the clade subtending that branch (posterior probability>90%).

Accepted Manuscript

ACCEPTED MANUSCRIPT

Table 1. Demographic characteristics of the study population.

Characteristic	HPV Genotype	
	HPV 53	**HPV 66**
Patients, N	30	24
Age (years), Median (IQR)	30.0 (24.0-39.0)	29.2 (24.8-33.6)
Cytology, N (%)		
Negative	9 (30.0)	2 (8.3)
ASCUS	15 (50.0)	8 (33.3)
L-SIL*	6 (20.0)	14 (58.3)
Gene, N (%)		
L1	22 (73.3)	22 (91.7)
E6	27 (90.0)	23 (95.8)
E7	26 (86.7)	21 (87.5)
LCR	24 (80.0)	7 (29.2)

ASCUS, typical cells of undetermined significance; L-SIL, low-grade squamous intraepithelial lesions; LCR, long control region.
* L-SIL lesions were statistically more frequently detected in patients infected with HPV 66 than in patients with HPV 53, P=0.005 by Fisher Exact Test.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Table2. Genetic variability of HPV 53 sequences.

A		L1Region[a]												
WT Codon		ACT	**TCC**		TCT	**TAT**	AGA	GTG	**CAA**	CCC	**CCT**		**GAC**	**AAT**	TTG
Nucleotide Position Interested[b]		6677	**6682-3**	6683	6704	**6733-4**	6740	6908	**6932**	6942	**6949**	**6949-50**	**6962**	**6993**	6998
Mutated Codon		ACA	**TGT**	TCT	TCC	**TTA**	AGG	GTA	**CAG**	TCC	**CTT**	**CTG**	**GAT**	**GAT**	TTA
WT Amino Acid		T	**S**		S	**Y**	R	V	**Q**	P	**P**		**D**	**N**	L
Amino AcidicPosition[c]		341	**343**		350	**360**	362	418	**426**	430	**432**		**436**	**447**	448
Mutated Amino Acid[d]		-	**C**	-	-	**L**	-	-	**-**	S	**L**	**L**	**-**	**D**	-
Cytology, N(%)	**Negative, N=6**	2 (33.3)	1 (16.7)	1 (16.7)	2 (33.3)	0 (0.0)	2 (33.3)	2 (33.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (33.3)	0 (0.0)	2 (33.3)
	ASCUS, N=11	1 (9.1)	0 (0.0)	5 (45.5)	5 (45.5)	0 (0.0)	5 (45.5)	5 (45.5)	0 (0.0)	4 (36.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	5 (45.5)
	L-SIL, N=5	2 (40.0)	0 (0.0)	4 (80.0)	4 (80.0)	1 (20.0)	4 (80.0)	4 (80.0)	1 (20.0)	1 (20.0)	1 (20.0)	1 (20.0)	0 (0.0)	1 (20.0)	2 (40.0)
Overall, N(%), n=22		5 (22.7)	1 (4.5)	10 (45.5)	11 (50)	1 (4.5)	11 (50)	11 (50)	1 (4.5)	5 (22.7)	1 (4.5)	1 (4.5)	2 (9.1)	1 (4.5)	9 (40.9)

B		E7 Region		
WT Codon		CTG	TCA	ATT
Nucleotide Position Interested[b]		695	803	823
Mutated Codon		TTG	CCA	ATG
WT Amino Acid		L	S	I
Amino AcidicPosition[c]		43	79	85
Mutated Amino Acid[d]		-	P	M
Cytology, N(%)	**Negative, N=7**	3 (42.9)	2 (28.6)	3 (42.9)
	ASCUS, N=13	2 (15.4)	1 (7.7)	2 (15.4)
	L-SIL, N=6	2 (33.3)	2 (33.3)	3 (50.0)
Overall, N(%), n=26		7 (26.9)	5 (19.2)	8 (30.8)

C		E6 Region[a]																	
WT Codon		CGT	CAG	GAG		CAT	CTA	AAA		TTC	GCT	ACA	TTA	AAA	TAT	TAC	GCA	TCA	GAC	**GCA**
Nucleotide Position Interested[b]		110	112	129	131	152	158	174	175	209	230	256	261	305	345	350	382	412	473	**546**
Mutated Codon		CGC	CGG	CAG	GAA	CAC	CTC	CAA	ACA	TTT	GCG	AAA	CTA	AAG	CAT	TAT	GAA	TTA	GAG	**TCA**
WT Amino Acid		R	Q	E		H	L	K		F	A	T	L	K	Y	Y	A	S	D	**A**
Amino AcidicPosition[c]		3	4	10		17	19	25		36	43	52	54	68	82	83	94	104	124	**149**
Mutated Amino Acid[d]		-	R	Q	-	-	-	Q	T	-	-	K	-	-	H	-	E	L	E	**S**
Cytology, N(%)	**Negative, N=8**	0 (0.0)	2 (25.0)	3 (37.5)	5 (62.5)	5 (62.5)	3 (37.5)	2 (25.0)	0 (0.0)	1 (12.5)	5 (62.5)	3 (37.5)	5 (62.5)	0 (0.0)	3 (37.5)	3 (37.5)	3 (37.5)	2 (25.0)	0 (0.0)	1 (12.5)
	ASCUS, N=13	3 (23.1)	2 (15.4)	2 (15.4)	6 (46.2)	6 (46.2)	2 (15.4)	1 (7.7)	1 (7.7)	0 (0)	6 (46.2)	2 (15.4)	6 (46.2)	5 (38.5)	2 (15.4)	3 (23.1)	4 (30.8)	1 (7.7)	1 (7.7)	1 (7.7)
	L-SIL, N=6	1 (16.7)	3 (50.0)	3 (50.0)	2 (33.3)	2 (33.3)	3 (50.0)	2 (33.3)	1 (16.7)	1 (16.7)	2 (33.3)	3 (50.0)	2 (33.3)	1 (16.7)	2 (33.3)	2 (33.3)	2 (33.3)	3 (50.0)	0 (0.0)	0 (0.0)
Overall, N(%), n=27		4 (14.8)	7 (25.9)	8 (29.6)	13 (48.1)	13 (48.1)	8 (29.6)	5 (18.5)	2 (7.4)	2 (7.4)	13 (48.1)	8 (29.6)	13 (48.1)	6 (22.2)	7 (25.9)	8 (29.6)	9 (33.3)	6 (22.2)	1 (3.7)	2 (7.4)

D		Long Control Region																
Wild Type Nucleotide		T	C	C	G	T	G	T	T	T	A	C	C	C	G	C	A	C
Position[b]		7422	7428	7432	7508	7528	7531	7618	7619	7634	7655	7676	7685	7756	7758	7784	7797	7810
Mutated Nucleotide		A	T	G	A	C	T	G	C	C	G	G	T	T	A	T	C	T
Cytology, N(%)	**Negative, N=9**	3 (33.3)	5 (55.6)	3 (33.3)	6 (66.7)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)	3 (33.3)
	ASCUS, N=15	5 (33.3)	6 (40.0)	2 (13.3)	11 (73.3)	5 (33.3)	5 (33.3)	5 (33.3)	5 (33.3)	6 (40.0)	5 (33.3)	5 (33.3)	5 (33.3)	5 (33.3)	5 (33.3)	5 (33.3)	5 (33.3)	2 (13.3)
	L-SIL, N=6	4 (66.7)	4 (66.7)	2 (33.3)	6 (100)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	4 (66.7)	3 (50.0)
Overall N(%), N=30		12 (40.0)	15 (50.0)	7 (23.3)	23 (76.7)	12 (40.0)	12 (40.0)	12 (40.0)	12 (40.0)	13 (43.3)	12 (40.0)	12 (40.0)	12 (40.0)	12 (40.0)	12 (40.0)	12 (40.0)	12 (40.0)	8 (26.7)

[a]Novel HPV 53 mutations are reported in bold.

[b] Nucleotides were numbered according to the HPV 53 X74482 reference sequence, considering the whole HPV 53 genome.

[c] Amino acids were numbered starting from the first amino acid of the specific protein.

[d] Synonymous changes are reported with a dash "-".

Accepted Manuscript

ACCEPTED MANUSCRIPT

WT, wild type; ASCUS, atypical cells of undetermined significance; L-SIL, low-grade squamous intraepithelial lesions.

Accepted Manuscript

ACCEPTED MANUSCRIPT

Table3. Genetic variability of HPV 66 sequences

A	L1 Region[a]											
WT Codon	AGC	GCC	AAT	**GGC**		TTA	**TCC**	CAG	**Ins**	GAA	**CTG**	GAT
Nucleotide Position Interested[b]	6660	6711	6849	6855	**6855**	6858	**6859**	6927	**6954**	6984	**7016**	7017
Mutated Codon	AGT	GCA	AAC	GGA	**GGT**	TTG	**CCC**	CAA	**CAT**	GAG	**CTA**	AAT
WT Amino Acid	S	A	N	**G**		L	**S**	Q	**Ins**	E	**L**	D
Amino AcidPosition[c]	338	355	401	**403**		404	**405**	427	**436-437**	446	**457**	458
Mutated Amino Acid[d]	-	-	-	-	-	-	**P**	-	**H**	-	-	N
Cytology, N(%) Negative, n=2	1 (50.0)	1 (50.0)	1 (50.0)	2 (100)	0 (0.0)	1 (50.0)	0 (0.0)	1 (50.0)	1 (50.0)	1 (50.0)	1 (50.0)	0 (0.0)
ASCUS, n=7	4 (57.1)	4 (57.1)	4 (57.1)	4 (57.1)	1 (14.3)	4 (57.1)	0 (0.0)	4 (57.1)	0 (0.0)	4 (57.1)	4 (57.1)	0 (0.0)
L-SIL, n=13	11 (84.6)	11 (84.6)	11 (84.6)	11 (84.6)	2 (15.4)	11 (84.6)	2 (15.4)	11 (84.6)	0 (0.0)	11 (84.6)	7 (53.8)	2 (15.4)
Overall, N(%), n=22	16 (72.7)	16 (72.7)	16 (72.7)	17 (77.3)	3 (13.6)	16 (72.7)	2 (9.1)	16 (72.7)	1 (4.5)	16 (72.7)	12 (54.5)	2 (9.1)

B	E7 Region[a]					
WT Codon	**GTA**	**GTT**	**ACG**	**TTG**	**GCA**	**TCT**
Nucleotide Position Interested[b]	**585**	**611**	**637**	**788-790**	**886**	**890**
Mutated Codon	**GCA**	**ATT**	**ACT**	**CTT**	**GCG**	**ACT**
WT Amino Acid	**V**	**V**	**T**	**L**	**A**	**S**
Amino AcidPosition[c]	**5**	**11**	**19**	**70**	**102**	**104**
Mutated Amino Acid[d]	**A**	**I**	-	-	-	**T**
Cytology, N(%) Negative, n=2	1 (50.0)	1 (50.0)	1 (50.0)	1 (50.0)	1 (50.0)	1 (50.0)
ASCUS, n=7	5 (71.4)	5 (71.4)	5 (71.4)	5 (71.4)	3 (42.9)	4 (57.1)
L-SIL, n=12	11 (91.7)	11 (91.7)	11 (91.7)	11 (91.7)	11 (91.7)	11 (91.7)
Overall, N(%), n=21	17 (81.0)	17 (81.0)	17 (81.0)	17 (81.0)	15 (71.4)	16 (76.2)

C	E6 Region								
WT Codon	CTA	CTA	**TAT**	**TAT**	AGT	**GAG**	**ACG**		GCT
Nucleotidic Position Interested	234	264	**338**	**360**	378	**449**	**539**	539	551
Mutated Codon	TTA	TTA	**TAC**	**GAT**	GGT	**GAA**	**ACA**	ACC	GCC
WT Amino Acid	L	L	**Y**	**Y**	S	**E**	**T**		A
Amino Acidic Position	45	55	**79**	**87**	93	**116**	**146**		150
Mutated Amino Acid	-	-	-	**D**	G	-	-	-	-
Cytology, N(%) Negative, n=2	2 (100)	1 (50.0)	1 (50.0)	0 (0.0)	1 (50.0)	0 (0.0)	1 (50.0)	1 (50.0)	1 (50.0)
ASCUS, n=8	5 (62.5)	4 (50.0)	1 (12.5)	1 (12.5)	4 (50.0)	2 (25.0)	0 (0.0)	2 (25.0)	4 (50.0)
L-SIL, n=14	12 (85.7)	12 (85.7)	2 (14.3)	1 (7.1)	12 (85.7)	1 (7.1)	0 (0.0)	7 (50.0)	12 (85.7)
Overall, N(%), n=24	19 (79.2)	17 (70.8)	4 (16.7)	2 (8.3)	17 (70.8)	3 (12.5)	1 (4.2)	10 (41.7)	17 (70.8)

D	Long Control Region									
WT Codon	**A**	**A**	**C**	**T**	**A**	**T**	**A**	**G**	**G**	**A**
Nucleotidic Position Interested	**7397**	**7426**	**7438**	**7530**	**7539**	**7620**	**7739**	**7743**	**7748**	**7779**
Mutated Codon	**C**	**G**	**T**	**G**	**G**	**A**	**T**	**T**	**T**	**T**
Cytology, N(%) Negative, n=2	1 (50.0)	0 (0.0)	1 (50.0)	0 (0.0)	0 (0.0)	2 (100)	0 (0.0)	0 (0.0)	0 (0.0)	1 (50.0)
ASCUS, n=2	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (50.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
L-SIL, n=3	1 (33.3)	1 (33.3)	2 (66.7)	1 (33.3)	0 (0.0)	2 (66.7)	1 (33.3)	1 (33.3)	1 (33.3)	2 (66.7)
Overall, N(%), n=7	2 (28.6)	1 (14.3)	3 (42.9)	1 (14.3)	1 (14.3)	4 (57.1)	1 (14.3)	1 (14.3)	1 (14.3)	3 (42.9)

[a]Novel HPV 66 mutations are reported in bold.

[b] Nucleotides were numbered according to the HPV 66 reference sequence U31794, considering the whole HPV 66 genome.

[c] Amino acids were numbered starting from the first amino acid of the specific protein.

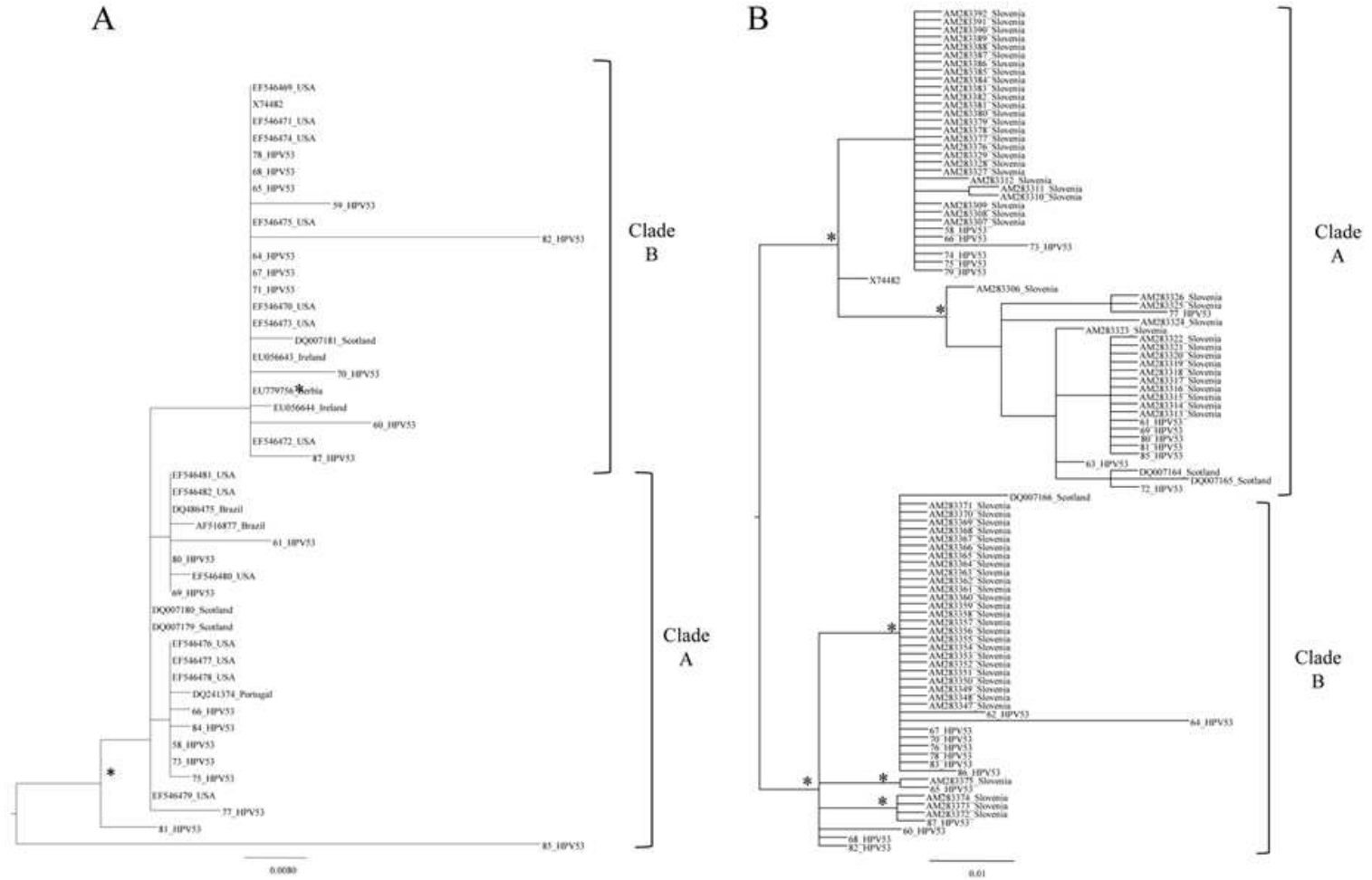
[d] Synonymous changes are reported with a dash "-".

WT, wild type; ASCUS, atypical cells of undetermined significance; L-SIL, low-grade squamous intraepithelial lesions.

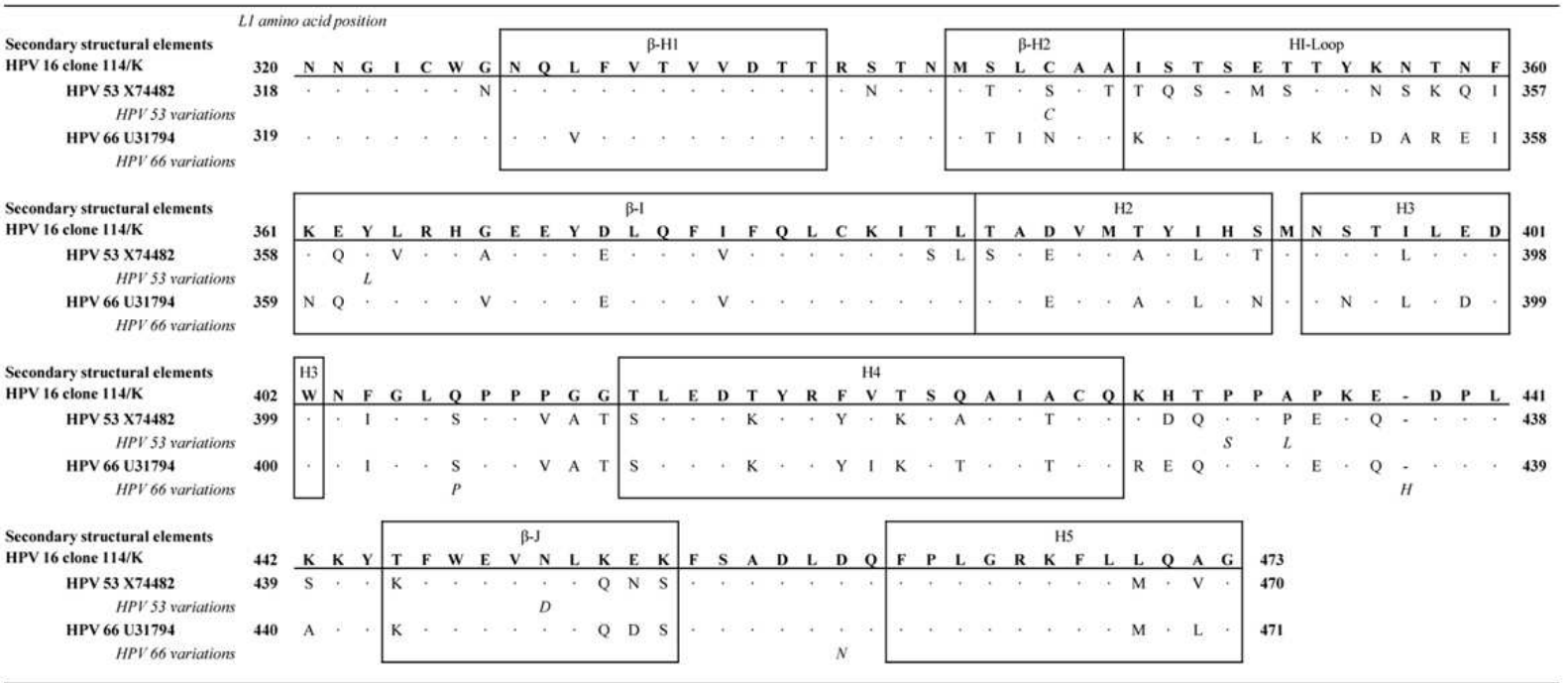
Accepted Manuscript

ACCEPTED MANUSCRIPT

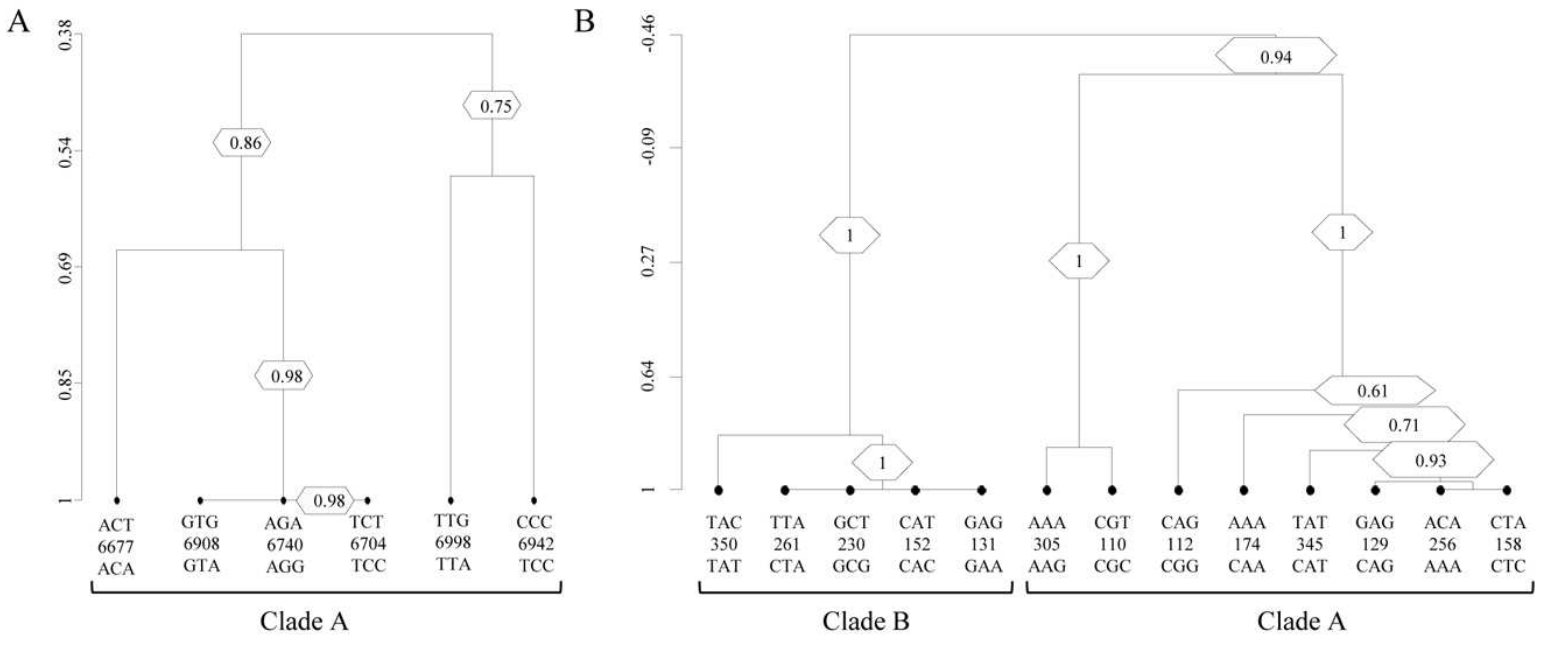
Manuscript



ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT

Manuscript