RESEARCH COMMUNICATION

Whole Genome Analysis of Human Papillomavirus Type 16 Multiple Infection in Cervical Cancer Patients

Jira Chansaenroj[1], Apiradee Theamboonlers[1], Pairoj Junyangdikul[2], Sukumarn Swangvaree[3], Anant Karalak[4], Yong Poovorawan[1]*

Abstract

The characterization of the whole genome of human papillomavirus type 16 (HPV16) from cervical cancer specimens with multiple infections in comparison with single infection samples as the oncogenic potential of the virus may differ. Cervical carcinoma specimens positive for HPV16 by PCR and INNO-LiPA were randomly selected for whole genome characterization. Two HPV16 single infection and six HPV16 multiple infection specimens were subjected to whole genome analysis by using conserved primers and subsequent sequencing. All HPV16 whole genomes from single infection samples clustered in the European (E) lineage while all multiple infection specimens belonged to the non-European lineage. The variations in nucleotide sequences in E6, E7, E2, L1 and Long control region (LCR) were evaluated. In the E6 region, amino acid changes at L83V were related to increased cancer progression. An amino acid variation N29S within the E7 oncoprotein significantly associated with severity of lesion was also discovered. In all three domains of the E2 gene non synonymous mutations were found. The L1 region showed various mutations which may be related to conformation changes of viral epitopes. Some transcription factor binding sites in the LCR region correlated to virulence were shown on GRE/1, TEF-1, YY14 and Oct-1. HPV16 European variant prone to single infection may harbor a major variation at L83V which significantly increases the risk for developing cervical carcinoma. HPV16 non-European variants prone to multiple infections may require many polymorphisms to enhance the risk of cervical cancer development.

Keywords: Human papillomavirus - whole genome - multiple infections - Thai cervical cancer patients

Introduction

Worldwide, high-risk HPV types (HR-HPV) are the cause of cervical cancer. HPV16 is the most prevalent HPV type found in cervical cancer patients (Bosch et al., 1995). The HPV genome comprises approximately 7.9 kb of circular double-stranded DNA which consists of two genes, the early (E) and late (L) region, respectively. The early region includes the open reading frames E1, E2, E4, E5, E6 and E7 which are associated with multiple functions of replication, transcription, transformation and viral adaptation; the late region which covers almost 40% of the genome, is translated into the L1 and L2 capsid proteins. The functions of these two proteins are viral DNA packaging and maturation. The remaining ones are first, the Upstream Regulatory Region (URR) of 1,000 bp containing cis-elements for the control of viral replication and transcription. The second one is a non-coding region (NCR) located between E5 and L2 which represents the most highly variable region (Muñoz et al., 2006; Burk et al., 2009; Smith et al., 2011).

The worldwide prevalence of invasive cervical cancer amounts to 52-58 % caused by HPV16, 13-22 % by HPV18 and the remaining percent are due to other genotypes depending on geographic variations (Smith et al., 2007). The main etiological factors of cervical cancer development may depend on persistence of HR-HPV infection and the variants of HPV (Pillai et al., 2009, Zuna et al., 2009). HPV variants are described as viral nucleotide sequences that vary in specific regions of the genome with less than 2% variation in the coding region and less than 5% in the non-coding regions (Bernard et al., 1994). Thus, the whole genome of HPV16 and its variants in E6, E7, E2, E5, L1, L2 and Long Control Region (LCR) help us understand the molecular basis for its oncogenic potential and viral persistence (Yamada et al., 1997). HPV16 variants were found in different biological environments and geographical locations that coevolved in the three ethnic groups of Africans, Asians and Caucasians (Icenogle et al., 1991). From those three populations, the variants of Asian (As), Asian-American (AA), African 1 (Af-1), African 2 (Af-2) and European (E) and a recent variant of North American 1 (NA1) HPV have originated. These five variants have been identified based

[1]*Center of Excellence in Clinical Virology, Department of Paediatrics, Faculty of Medicine, Chulalongkorn University,* [2]*Department of Pathology, Samitivej Srinakharin Hospital,* [3]*Department of Gynecologic Oncology,* [4]*Department of Pathology, National Cancer Institute, Bangkok, Thailand *For correspondence: yong.p@chula.ac.th*

on nucleotide changes in E6, L1 and LCR. The European variants are usually found in all regions except Africa. The Asian variants are a subclass of the European lineage and commonly detected in South-East Asians (Yamada et al., 1997). One of the potential risk factors for persistence and progression of HPV infection may involve variants of HPV types. Studies on multiethnic populations have also shown other non-European variants of HPV 16 and 18 associated with the development of cervical lesions, increased risk of persistence and immunogenicity (Villa et al., 2000). These variants may express differences in transcriptional regulation, biological activity or immune response to specific viral epitopes (Veress et al., 2001; Kämmer et al., 2002). The non-European variants of HPV16 may pose a two-to nine-fold increased risk of HSIL and cervical cancer depending on the respective populations (Hildesheim et al., 2002). The diversity in oncogenic potential of HPV16 and 18 is due to geographic variations and prevalence of the predominant molecular variants (Sichero et al., 2007).

HPV16 of the European variant is most widely circulated worldwide except in Africa. Various studies on LCR, L1, L2 and E6 genes of HPV16 variants showed rare or nonexistent recombination between variants (Yamada T et al., 1995). Asian-American variants of HPV16 stimulate p53 degradation and also increase p97 promoter activity. These may increase viral oncogene expression and reveal some functional difference when compared with the HPV16 prototype (Stöppler et al., 1996; Veress et al., 1999). The E6 oncoprotein interacting with cellular proteins could lead to cell transformation to cancer by altering the control of cell cycles and apoptosis with simultaneous degradation of the cellular tumor suppressor protein p53. The E6 variants regulate tumorigenesis by the Notch signaling pathway and oncogenic Ras differentiation suggesting that they could feasibly enhance the oncogenic potential of the Asian American variants (Chakrabarti et al., 2004). Persistence of HPV16 and progression to cervical intraepithelial neoplasia grade 2/3 may be associated with variants exhibiting a non-synonymous substitution at nucleotide position T350G (Londesborough et al., 1996). The E7 oncoprotein inactivates the retinoblastoma gene product (pRb) which leads to uncoupling of centrosome duplication and loss of control of the cell division cycle. The E2 viral protein regulates transcription by binding adjacent to the promoter. Increased E2 expression represses E6/E7 (oncogene) expression. Destruction of the E2 gene induces E6/E7 oncogene transcription during viral integration in the viral genome prototype whereas, in Asian-American variants, the E2 gene is intact which may account for higher replication efficiency in comparison with European variants (Tan et al., 1994; Jeon et al., 1995). The L1 protein constitutes 83% of the capsid and assists in the assembly of virus-like particles (VLPs) in well-differentiated cells of the superficial layers of the cervical epithelium used as the target for the design of prophylactic HPV vaccines. The infectious potential of different variants and defining epitopes relevant for vaccine design are important for discrimination (Sun et al., 2011).

The non-synonymous variations in L1 epitopes play a role in mediating escape from neutralization antibody response (Frati et al., 2011). The LCR regulates replication and transcription. It has been divided into three functionally distinct segments, i.e., the 5'segment, the central and the 3' segment. The 5'segment contains a negative regulatory element and nuclear matrix attachment region (MAR) which increases mRNA stability and represses viral oncoprotein expression. The central segment (NCR region) serves as an epithelial-specific transcription enhancer. The HPV16 enhancer harbors more than 20 binding sites for multiple transcription factors such as transcriptional enhancer factor-1 (TEF-1) and glucocorticoid responsive element (GRE) (Durst et al., 1992; Sibbet et al., 1995; Tan et al., 1998; Stunkel et al., 1999). The 3' region includes the origin of virus replication and the E6, E7 promoter. The AA variants displayed a 3-fold increase in p97 promoter activity compared to the E variants. Therefore, the sequence variations within LCR, E6 and E7 genes may have an effect on oncogenesis. However, other factors such as host factors may be associated with the activation of oncogenicity, as nucleotide changes in variant genes can induce changes in other genes (Casas et al., 1999; Lizano M et al., 2009).

In conclusion, the polymorphisms in the genome of HPV variants have an effect on oncogenicity. The aim of the present study has been to analyze variants of HPV 16 and characterize the whole genomes of cervical cancer specimens with multiple infections.

Materials and Methods

This protocol was approved by the Ethics Committee of the hospital and Faculty of Medicine, Chulalongkorn University (IRB148/53). The 330 fresh tissues diagnosed as cervical cancer by cytology and histology were obtained from the National Cancer Institute (NCI) suspended in 2 ml each of phosphate buffered saline from December, 2010 until end of September, 2011. The specimens were stored in liquid nitrogen until used. HPV DNA was detected by using consensus primers for PCR amplification and further analyzed by direct sequencing. Whole genome analysis was performed in samples with multiple infections with HPV16 and other genotypes in comparison with the whole genome of HPV16 single infection.

DNA extraction

DNA was extracted applying the standard organic method (phenol-chloroform) and alcohol precipitation of the specimens as described by Broccolo F et al. (Broccolo F et al., 2005). The purified material was re-suspended in a final volume of 30 ul of deionized water.

HPV detection and typing

HPV DNA was detected by using consensus polymerase chain reaction of the E1 and L1 regions based on a previous study (Lurchachaiwong et al., 2009). Positive and negative controls were included in

each assay. HPV DNA positive samples were identified by electrophoresis in 2% agarose gel (FMC Bioproducts Rockland, ME). The HPV DNA was further purified with the agarose gel extract mini kit (5PRIME, Darmstadt, Germany) according to the manufacturer's specifications. The purified DNA was sequenced by FirstBASE Laboratories SDNBHD (Selangor Darul Ehsan, Malaysia). The nucleotide sequences were analyzed by Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm. nih.gov/Blast.cgi).

Of those samples amplified by PCR, 158 HPV positive samples were randomly selected to detect multiple infections applying the INNO-LiPA method (INNO-LiPA Genotyping Extra, Innogenetics, Belgium). The typing results were interpreted according to the manufacturer's instructions.

HPV16 Whole genome sequencing and phylogenetic tree construction

The multiple infection specimens that co-infected with HPV16 and the HPV16 single infection specimens were selected for whole genome characterization. Conserved primers were used for whole genome amplification and sequencing as previously described (Lurchachaiwong et al., 2009). The Lasergene 6 Package (DNASTAR, Inc., Madison, WI) was used to assemble the nucleotide sequences. The completed genome sequences were compared with the reference sequences available at GenBank. The alignments were performed by using Clustal W applying the BioEdit program (version 7.0.4.1) (Hall T.A., 1999). The phylogenetic trees were constructed for whole genome sequences and for each gene by neighbor-joining (NJ) from MEGA 4.0 program (Tamura et al., 2007) and support tree topologies by Bootstrapping applied with 1,000 replicates. In this study, the whole genomes of six HPV16 multiple infections have been analyzed with the two whole genomes of HPV16 single infection (Table 1). The base positions in the published sequence HPV16R (Myers et al., 1997) were used for analysis.

Results

Of the 330 samples, 293 (88.8%) were positive for HPV DNA in the E1 and L1 regions. The 158 samples of HPV positive specimens which were tested by INNO-LiPA method showed 10 (6.3%) co-infections. The highest incidence of co-infection in this study showed five different genotypes in the same sample.

Whole genome sequence comparison analyses of HPV16 multiple infections and HPV16 single infections are shown in Tables 1 and 2. As can be seen in Table 1, the nucleotide sequences in E6, E7, E2, L1 and non-coding region at positions 7155-7905 (LCR) showed various variation points. Most variations were found in multiple infection groups, especially the Asian-American variant group. The LCR region showed much more variations than other regions. Twenty-three positions of nucleotide variation were detected in the LCR in contrast to only five positions in the E7 gene. The amino acids encoded by all four genes, E6, E7, E2 and L1 were altered as indicated

in Table 2.

The E6 gene of the 8 whole genome sequences showed 7 mutation points; 3 (42.9%) were synonymous mutations while 4 (57.1%) were non synonymous mutations with amino acid substitutions at Q14H (nt G145T), D25E (nt T178G), H78Y (nt C335T), L83V (nt T350G) (Table 2). In this study, the E7 gene was more conserved than the E6 gene when compared with the reference strains (Table 1). Four synonymous mutations were found in this region. One non- synonymous mutation at N29S (nt A647G) confined to the East Asian group was found in this study. The two whole genome samples with single infection were related to European variants.

The E2 gene of HPV16 single infection displayed 7 (26.9%) synonymous mutations and 19 (73.1%) non-synonymous mutations with amino acid substitutions as shown in Table 2. All sequence variations were found in multiple infection groups while two point mutations at nt A2926G and C3410T were found in all groups.

The L1 gene of HPV16 harbored 10 (58.8%) synonymous mutations and 7 (41.2%) non- synonymous mutations resulting in amino acid substitutions at as shown in Table 2.

Table 1. Nucleotide Sequences Analysis Among the HPV16 E6, E7, E2, L1 and LCR Variants Compared to the Reference Sequence

	E6						
Position	145	178	286	289	335	350	532
Reference (HPV16R)	G	T	T	A	C	T	A
K02718 (E)							
AF472508 (Af-1)	T		A	G	T		
AF472509 (Af-2)	T		A	G	T		
AF534061 (EA)		G					
AF402678 (AA)	T		A	G	T	G	G
FJ610152							
JQ004093						G	
JQ004092						G	
JQ004094 (HPV16+33+69/71)		G					
JQ004096 (HPV16+33+58+68+69/71)		G					
JQ004099 (HPV16+52)		G					
JQ004097 (HPV16+33+39+58)	T		A	G	T	G	G
JQ004095 (HPV16+31+39)	T		A	G	T	G	G
JQ004098 (HPV16+31+58)	T		A	G	T	G	G

	E7				
Position	647	732	789	795	846
Reference (HPV16R)	A	T	T	T	T
K02718 (E)					
AF472508 (Af-1)			C	G	
AF472509 (Af-2)	G		C	G	
AF534061 (EA)	G				C
AF402678 (AA)		C	C	G	
FJ610152					
JQ004093					
JQ004092					
JQ004094 (HPV16+33+69/71)		G			C
JQ004096 (HPV16+33+58+68+69/71)	G				C
JQ004099 (HPV16+52)		G			C
JQ004097 (HPV16+33+39+58)		C	C	G	
JQ004095 (HPV16+31+39)		C	C	G	
JQ004098 (HPV16+31+58)		C	C	G	

E2

Sample	2800 +28	2800 +60	2926	3100 +59	3100 +81	3100 +82	3200 +24	3200 +49	3200 +62	3300 +77	3300 +84	3300 +87	3400 +10	3400 +49	3500 +16	3500 +17	3500 +24	3500 +38	3500 +66	3600 +64	3600 +84	3600 +94	3700 +06	3700 +78	3700 +87	3805
Reference K02718 (E)	G	C	A	C	A	G	T	G	A	C	T	T	T	C	G	C	T	T	A	T	T	C	T	T	G	C
AF472508 (Af-1)	A	G	A	A			A	G	G			T	A	A					G	A	A		T	A		
AF472509 (Af-2)	A	G	A	A			A	G	G			T	A	A	C	C		G		A	A	C	T	A	G	
AF534061 (EA)	A		G	A			A		C				T	A	C					A				A		
AF402678 (AA)	A	G	A	C	A	A	A	G	G		C	T	A	A	C		C	G	C	A	A	C	T	A	G	
FJ610152		G							C				A							A						
JQ004093		G											T													
JQ004092		G											T													
JQ004094 (HPV16+33+69/71)	A		G	A			A		C				T	A	C					A				A		
JQ004096 (HPV16+33+58+68+69/71)	A		G	A			A		C				T	A	C					A				A		
JQ004099 (HPV16+52)	A		G	A			A		C				T	A	C					A				A		
JQ004097 (HPV16+33+39+58)	A	G	A	C	A	A	A	G	G		C	T	A	A	C		C	G	C	A	A	C	T	A	G	
JQ004095 (HPV16+31+39)	A	G	A	C	A	A	A	G	G		C	T	A	A	C		C	G	C	A	A	C	T	A	G	
JQ004098 (HPV16+31+58)	A	G	A	C	A	A	A	G	G		C	T	A	A	C		C	G	C	A	A	C	T	A	G	

L1

Position	5564	5698	5864	5911	5983	6165	6180	6247	6434	6559	6695	6721	6854	6865	6970	6994	7060
Reference (HPV16R)	C	G	C	T	G	C	A	T	A	C	A	G	C	C	C	G	G
AF472508 (Af-1)	G	A	T	C		A		C	G	T		A	T		T	A	T
AF472509 (Af-2)	G	A	T	C		A	C	C	G	T	C	A	T	T	T	A	T
AF534061 (EA)							C		G								A
AF402678 (AA)	G	A	T	C		A	C	C	G	T	C	A	T	T	T	A	T
FJ610152									G								
JQ004093									G								A
JQ004092									G								
JQ004094 (HPV16+33+69/71)									G								A
JQ004096 (HPV16+33+58+68+69/71)							C		G						T		A
JQ004099 (HPV16+52)									G								A
JQ004097 (HPV16+33+39+58)	G	A	T	C	A	A	C	C	G	T	C	A	T	T	T	A	T
JQ004095 (HPV16+31+39)	G	A	T	C	A	A	C	C	G	T	C	A	T	T	T	A	T
JQ004098 (HPV16+31+58)	G	A	T	C	A	A	C	C	G	T	C	A	T	T	T	A	T

LCR

| Sample	7100 +75	7100 +77	7100 +93	7200 +01	7200 +33	7200 +70	7200 +87	7200 +89	7300 +39	7300 +94	7300 +95	7400 +85	7400 +89	7521	7600 +69	7600 +89	7700 +29	7700 +30	7700 +43	7700 +64	7700 +86	7800 +42	7800 +86
Reference (HPV16R)	A	T	G	T	A	C	A	A	A	C	C	A	G	G	C	C	A	A	T	C	C	G	C
AF472508 (Af-1)			T										A	A	A					T	T		
AF472509 (Af-2)			T		C							C	A	A	T					T	T		
AF534061 (EA)	C	C	T	C		T	C	C							A				C				A
AF402678 (AA)			T		C				T	T	T	C	A		A	T	A	C	G	T	T		A
FJ610152			T												A								
JQ004093			T												A								
JQ004092			T												A								
JQ004094 (HPV16+33+69/71)	C	C	T	C		T	C	C							A				C				A
JQ004096 (HPV16+33+58+68+69/71)	C	C	T	C		T	C	C							A				C				A
JQ004099 (HPV16+52)	C	C	T	C		T	C	C							A				C				A
JQ004097 (HPV16+33+39+58)			T						T	T	T	C	A		A	T	A	C	G	T	T		G
JQ004095 (HPV16+31+39)			T		C				T	T	T	C	A		A	T	A	C	G	T	T		G
JQ004098 (HPV16+31+58)			T		C				T	T	T	C	A		A	T	A	C	G	T	T		G

* The Nucleotide positions and numbering are based on the HPV16 reference sequence (HPV16R) (Myers G et al., 1997).

Table 2. Variation in Nucleotides and Amino Acid Sequences Change of HPV16 E6, E7, E2 and L1 Variants Compared to the Reference Sequence.

Gene	Nucleotide sequence		Amino acid residue
	Number	Change	Change
E6	145	G--> T	Q --> H
	178	T -->G	D --> E
	335	C --> T	H --> Y
	350	T -->G	L --> V
E7	647	A --> G	N --> S
E2	2826	G-->A	D --> N
	2860	C-->A	H --> Q
	3159	C --> A	T --> K
	3181	A --> C	E --> D
	3182	G --> A	A -->T
	3224	T --> A	L --> I
	3249	G --> A	R --> Q
	3362	A --> G	N --> D
	3377	C --> G	P --> A
	3384	T --> C	I -->T
	3387	T --> C	I --> T
	3410	C --> T	S --> P
	3449	G -->A	E --> K
	3516, 3517	CT --> AC	T -->N
	3566	T --> G	F --> V
	3684	C --> A	T --> K
	3778	G -->T	W --> C
	3787	C --> A	D --> E
L1	5564	C --> G	Q --> E
	5864	C --> T	H --> Y
	6165	C --> A	T --> N
	6180	A --> C	N --> T
	6434	A --> G	T --> A
	6695	A --> C	T --> P
	7060	G -->T/A	L --> F

* The Nucleotide positions and numbering are based on the HPV16 reference sequence (HPV16R) (Myers et al., 1997).

The NCR at nucleotide positions 4102-4235 displayed profound sequence variations, insertions and deletions while nucleotide positions 1-83 were highly conserved (data not shown). All 23 point mutations in the LCR region were ubiquitous among the multiple infection samples. The sequence variations in the LCR region at G7193T and G7521A were observed in all groups. All whole genome sequences were submitted to the GenBank database under accession numbers JQ004092 – JQ004099

Based on the phylogenetic tree constructed upon whole genome analysis, the multiple infection groups were separated into 2 lineages. Six samples with multiple infections were associated with non-European (Asian-American and East Asian) variants while the two samples with single infection were associated with European variants (Figure 1). According to Supplementary figure S1-8, the phylogenetic tree of each gene resulted the separation nearly the whole genome phylogenetic tree which multiple infections were separated into Asian-American and East Asian variants and single infection were related with European variants.

Discussion

Seventy –three percent of invasive cervical cancer
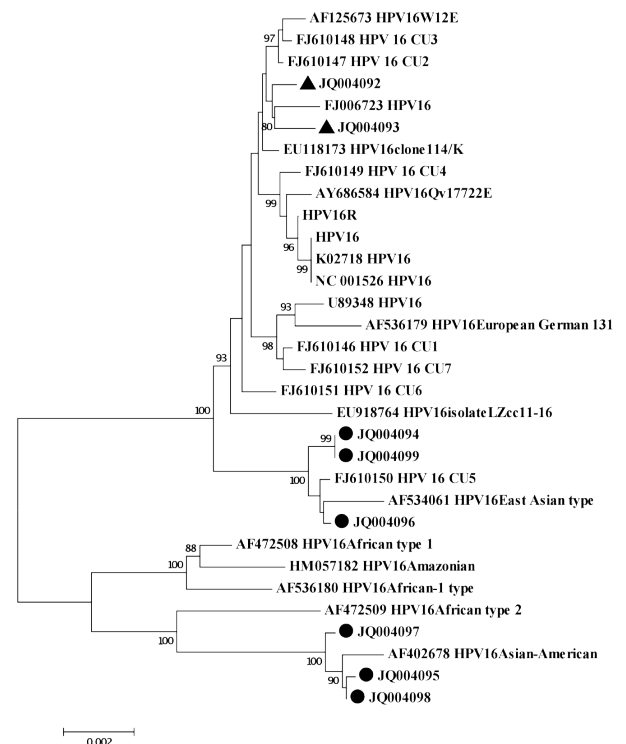
Figure 1. Phylogenetic Analysis of Whole Genome HPV16. ▲ = single infection, ● = multiple infection.

worldwide develops as a consequence of HPV16 and/ or HPV18 infection. Carcinogenesis is triggered by a combination of viral factors and non-viral factors. Normally, HPV16 is found in both single and multiple infections. Multiple HPV infections are associated with an increased risk of cervical precancerous lesions. Based on our study, 6.3% of HPV16 multiple infections in invasive cervical cancer were found by INNO-LiPA. Acquisition of multiple HPV types was associated with age because of behavioral and demographic variables (Spinillo et al., 2009). According to several studies, the variant types of HPV16 genome variations may be important to virus infectivity and pathogenicity (Hildesheim et al., 2001; Pista et al., 2007). The distribution of HPV variants did not depend on age, ethnicity or lifetime number of sexual partners (Xi et al., 1997). HPV16 variants have different biological and biochemical effects and hence, oncogenic potentials (Andersson S et al., 2000). Based on a previous report, HPV16 variants may show varying degrees of relation with progress of cervical neoplasia to cancer (Hildesheim et al., 2001). These variations may affect viral persistence, viral assembly, host immune responses and progression to invasive cervical cancer (Li et al., 2011; Stewart A.C et al., 1996). Hence, the identification of HPV genetic diversity in specific clinical settings may be important for the design of diagnostic and therapeutic modalities.

One important gene variation of HPV16 is found in the viral oncoprotein E6 gene. These variations result in amino acid changes which alter the immunogenic properties and biochemical activities as for example, induce p53 degradation, inhibit p53 transactivation, induce Bax degradation, binding to the E6-binding protein (E6BP) and the human discs large tumor suppressor protein (hDlg) (Mantovani et al., 2001). In this study, D25E and L83V

were detected which represent the major variation points that may be associated with the risk of cervical carcinoma. Both mutations occur due to genetic differences between populations. A change from Leucine to Valine in L83V which is part of oncoprotein E6, affects immunogenicity and induces cell transformation (Tornesello et al., 2000). The mutation D25E which has been reported as mainly distributed among Asian populations such as Chinese and Japanese populations, has also been discovered in this study (Shang Q et al., 2011). The mutation L83V, on the other hand, is identical to European and Asian-American variants (Yamada et al., 1997). Based on our results from phylogenetic tree analysis, the D25E mutation was found in the East Asian lineage while the L83V mutation was found in the Asian-American lineage (Figure 1). Both mutations may be associated with augmented viral persistence and progression to cancer.

The E7 sequence variations show geographic dependence. Almost none of the variations within the E7 gene affect the amino acid sequence or protein functions. According to a previous study, the amino acid changes at residues 37-49 were not significant for the transformation activity of the protein (Phelps et al., 1992). The only non-synonymous mutation N29S in E7 was found significantly more prevalent and related with the severity of oncogenic risk which would affect cell immortalization, enhance DNA integration in a manner dependent on geographic location (Song et al., 1997).

In the E2 region, we found nucleotide changes in all three domains; the transactivation domain at amino acid positions E142D (nt A3181C), A143T (nt G3182A), L157I (nt T3224A), R165Q (nt G3249A), hinge domain at amino acid positions T254N (nt C3516A, nt T3517C) and DNA-binding domain at amino acids positions F271V (nt T3566G), T310K (nt C3684A). The main disruption occurred in the E2 protein which includes the DNA-binding domain (nucleotides 3596-3872). Disruption of the C-terminal domain of E2 could possibly be related to the processing of promotion of viral persistence. In addition, amino acid residues 18-41 in the E2 region utilize the association within E1 to promote viral replication (Bhattacharjee et al., 2006). Some studies have pointed out that this variation along with C3684A could be related to the conformational alterations of DNA structure. However, the sequence variations in the E2 gene may not be the major mechanism responsible for enhancing the expression of E6 and E7 oncoproteins (Watts et al., 2001).

The L1 protein has five hypervariable immunodominant regions (BC, DE, EF, FG, HI loops) located within surface-exposed loops. The non-synonymous variations of the L1 gene which had occurred closely to neutralizing epitopes could lead to epitope conformation changes. This can be relevant to viral neutralization in that it may affect the efficiency of viral assembly into VLPs that escape the dendritic cell-dependent innate immunity in cervical cancer (Pastrana et al., 2001). The amino acid residues 83-97 affect the level of expression of the L1 protein. The variations observed in this study may be important to recognize the infection potential of different variants. This might be crucial to determine the epitopes relevant to vaccine development strategies and

therapeutic interventions (Sun et al., 2011). The novel five mutation points in the L1 gene at Q2E (nt C5564G), H102Y (nt C5864T), T202N (nt C6165A), T292A (nt A6434G) and T379P (nt A6695C) are located closely to the immunodominant epitope loops and thus, may affect viral antigenicity.

The LCR gene is the binding site of many cellular and viral transcription factors along with E2-binding sites. Based on previous functional studies, variations in the LCR contribute to the transformation capacity of high-risk HPV and induce enhancer activity while the sequences rearrangements can be involved in expression modifications. The expression level of each protein as well as the polymorphisms in their binding sites determines the threshold for virulence. In the present study, the LCR region showed variations of transcription factor binding sites in GRE/1 (at base positions C7394T and C7395T), TEF-1 (at base positions C7689A, T7743G), YY14 (at base positions A7729C, C7764T, C7786T) and Oct-1 (at base position G7842A) which may be associated with a risk for disease progression. Furthermore, approximately 80% of the variations of transcription factor sites at nucleotides G7193T and G7521A are normally found on a global scale and have no significant impact on disease progression. Whether wild-type or mutated, both sites play a role as integral parts of transcription factor binding sites and thus, influence the infection properties (Tornesello et al., 2000; Bhattacharjee et al., 2006; Kurvinen et al., 2000).

In summary, HPV variants are related to severity of lesions. Sequence variations in HPV16 probably enhance viral oncogene expression, immune evasion and thus, represent risk factors for developing cervical cancer. The E6 T350G variant may be associated with enhancing the oncogenic potential of HPV16 especially in European variants. On the other hand, HPV16 non-European variants, which are mainly found in multiple infection groups, may acquire polymorphisms in many viral genes that influence oncogenicity. The biological role of these variations is still unknown but they can help us determine the oncogenic potential and the effect of polymorphisms on host response. Research conducted on larger samples sizes of HPV16 variants may assist in elucidating these phenomena.

Acknowledgements

References

Andersson S, Alemi M, Rylander E, et al (2000). Uneven distribution of HPV 16 E6 prototype and variant (L83V) oncoprotein in cervical neoplastic lesions. *Br J Cancer*, **83**, 307-10.

Bernard HU, Chan SY, Manos MM, et al (1994). Identification and assessment of known and novel human papillomaviruses by polymerase chain reaction amplification, restriction fragment length polymorphisms, nucleotide sequence, and phylogenetic algorithms. *J Infect Dis*, **170**, 1077-85.

Bhattacharjee B, Sengupta S (2006). HPV16 E2 gene disruption and polymorphisms of E2 and LCR: some significant associations with cervical cancer in Indian women. *Gynecol Oncol*, **100**, 372-8.

Bosch FX, Manos MM, Muñoz N, et al (1995). Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. *J Natl Cancer Inst*, **87**, 796-802.

Broccolo F, Drago F, Careddu AM, et al (2005). Additional evidence that pityriasis rosea is associated with reactivation of human herpesvirus-6 and -7. *J Invest Dermatol*, **124**, 1234-40.

Burk RD, Chen Z, Van Doorslaer K (2009). Human papillomaviruses: genetic basis of carcinogenicity. *Public Health Genomics*, **12**, 281-90.

Casas L, Galvan SC, Ordoñez RM, et al (1999). Asian-american variants of human papillomavirus type 16 have extensive mutations in the E2 gene and are highly amplified in cervical carcinomas. *Int J Cancer*, **83**, 449-55.

Chakrabarti O, Veeraraghavalu K, Tergaonkar V, et al (2004). Human papillomavirus type 16 E6 amino acid 83 variants enhance E6-mediated MAPK signaling and differentially regulate tumorigenesis by notch signaling and oncogenic Ras. *J Virol*, **78**, 5934-45.

Durst M, Glitz D, Schneider A, zur Hausen H (1992). Human papillomavirus type 16 (HPV 16) gene expression and DNA replication in cervical neoplasia: analysis by in situ hybridization. *Virology*, **189**, 132-40.

Frati E, Bianchi S, Colzani D, et al (2011). Genetic variability in the major capsid L1 protein of human papillomavirus type 16 (HPV-16) and 18 (HPV-18). *Infect Genet Evol*, **11**, 2119-24.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*, **41**, 95-98.

Hildesheim A, Schiffman M, Bromley C, et al (2001). Human papillomavirus type 16 variants and risk of cervical cancer. *J Natl Cancer Inst*, **93**, 315-8.

Hildesheim A, Wang SS (2002). Host and viral genetics and risk of cervical cancer: a review. *Virus Res*, **89**, 229-40.

Icenogle JP, Sathya P, Miller DL, Tucker RA, Rawls WE (1991). Nucleotide and amino acid sequence variation in the L1 and E7 open reading frames of human papillomavirus type 6 and type 16. *Virology*, **184**, 101-7.

Jeon S, Allen-Hoffmann BL, Lambert PF (1995). Integration of human papillomavirus type 16 into the human genome correlates with selective growth advantage of cells. *J Virol*, **69**, 2989-97.

Kämmer C, Tommasino M, Syrjänen S, et al (2002). Variants of the long control region and the E6 oncogene in European human papillomavirus type 16 isolates: implications for cervical disease. *Br J Cancer*, **86**, 269-73.

Kurvinen K, Yliskoski M, Saarikoski S, Syrjänen K, Syrjänen S (2000). Variants of the long control region of human papillomavirus type 16. *Eur J Cancer*, **36**, 1402-10.

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

Lizano M, Berumen J, García-Carrancá A (2009). HPV-related carcinogenesis: basic concepts, viral types and variants. *Arch Med Res*, **40**, 428-34.

Londesborough P, Ho L, Terry G, et al (1996). Human papillomavirus genotype as a predictor of persistence and development of high-grade lesions in women with minor cervical abnormalities. *Int J Cancer*, **69**, 364-8.

Lurchachaiwong W, Junyangdikul P, Payungporn S, et al (2009). Relationship between hybrid capture II ratios and DNA amplification of E1, E6 and L1 genes used for the detection of human papillomavirus in samples with different cytological findings. *Asian Pac J Allergy Immunol*, **27**, 217-24.

Lurchachaiwong W, Junyangdikul P, Payungporn S, et al (2009). Entire genome characterization of human papillomavirus type 16 from infected Thai women with different cytological findings. *Virus Genes*, **39**, 30-8.

Mantovani F., Banks L (2001). The human papillomavirus E6 protein and its contribution to malignant progression. *Oncogene*, **20**, 7874-87.

Muñoz N, Castellsagué X, de González AB, Gissmann L (2006). Chapter 1: HPV in the etiology of human cancer. *Vaccine*, **24**, 1-10.

Myers G, et al (1997). Human papillomaviruses 1997: a compilation and analysis of nucleic acid and amino acid sequences. Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM.

Pastrana DV, Vass WC, Lowy DR, Schiller JT (2001). NHPV16 VLP vaccine induces human antibodies that neutralize divergent variants of HPV16. *Virology*, **279**, 361-9.

Phelps WC, Münger K, Yee CL, Barnes JA, Howley PM (1992). Structure-function analysis of the human papillomavirus type 16 E7 oncoprotein. *J Virol*, **66**, 2418-27.

Pillai MR, Hariharan R, Babu JM, et al (2009). Molecular variants of HPV-16 associated with cervical cancer in Indian population. *Int J Cancer*, **125**, 91-103.

Pista A, Oliveira A, Barateiro A, et al (2007). Molecular variants of human papillomavirus type 16 and 18 and risk for cervical neoplasia in Portugal. *J Med Virol*, **79**, 1889-97.

Shang Q, Wang Y, Fang Y, et al (2011). Human papillomavirus type 16 variant analysis of E6, E7, and L1 genes and long control region in cervical carcinomas in patients in northeast China. *J Clin Microbiol*, **49**, 2656-63.

Sibbet GJ, Cuthill S, Campo MS (1995). The enhancer in the long control region of human papillomavirus type 16 is up regulated by PEF-1 and down regulated by Oct-1. *J Virol*, **69**, 4006-11.

Sichero L, Ferreira S, Trottier H, et al (2007). High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. *Int J Cancer*, **120**, 1763-8.

Smith B, Chen Z, Reimers L, et al (2011). Sequence Imputation of HPV16 Genomes for Genetic Association Studies. *PLoS One*, **6**, 21375-?.

Smith JS, Lindsay L, Hoots B, et al (2007). Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. *Int J Cancer*, **121**, 621-32.

Song YS, Kee SH, Kim JW, et al (1997). Major sequence variants in E7 gene of human papillomavirus type 16 from cervical cancerous and noncancerous lesions of Korean women. *Gynecol Oncol*, **66**, 275-81.

36. Spinillo A, Dal Bello B, Gardella B, et al (2009). Multiple human papillomavirus infection and high grade cervical intraepithelial neoplasia among women with cytological diagnosis of atypical squamous cells of undetermined

significance or low grade squamous intraepithelial lesions. *Gynecol Oncol*, **113**, 115-9.

Stewart AC, Eriksson AM, Manos MM, et al (1996). Intratype variation in 12 human papillomavirus types: a worldwide perspective. *J Virol*, **70**, 3127-36.

Stöppler MC, Ching K, Stöppler H, et al (1996). Natural variants of the human papillomavirus type 16 E6 protein differ in their abilities to alter keratinocyte differentiation and to induce p53 degradation. *J Virol*, **70**, 6987-93.

Stunkel W, Bernard HU (1999). The chromatin structure of the long control region of human papillomavirus type 16 represses viral oncoprotein expression. *J Virol*, **73**, 1918-30.

Sun Z, Ren G, Cui X, et al (2011). Genetic diversity of HPV-16 E6, E7, and L1 genes in women with cervical lesions in Liaoning Province, China. *Int J Gynecol Cancer*, **21**, 551-8.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*, **24**, 1596-99.

Tan SH, Leong LEC, Walker PA, Bernard HU (1994). The human papillomavirus type 16 E2 transcription factor binds with low cooperativity to two flanking sites and represses the E6 promoter through displacement of Sp1 and TFIID. *J Virol*, **68**, 6411-20.

Tan SH, Bartsch D, Schwarz E, Bernard HU (1998). Nuclear matrix attachment regions of human papillomavirus type 16 point toward conservation of these genomic elements in all genital papillomaviruses. *J Virol*, **72**, 3610-22.

Tornesello ML, Buonaguro FM, Buonaguro L, et al (2000). Identification and functional analysis of sequence rearrangements in the long control region of human papillomavirus type 16 Af-1 variants isolated from Ugandan penile carcinomas. *J Gen Virol*, **81**, 2969-82

Veress G, Szarka K, Dong XP, Gergely L, Pfister H (1999). Functional significance of sequence variation in the E2 gene and the long control region of human papillomavirus type 16. *J Gen Virol*, **80**, 1035-43.

Veress G, Murvai M, Szarka K, et al (2001). Transcriptional activity of human papillomavirus type 16 variants having deletions in the long control region. *Eur J Cancer*, **37**, 1946-52.

Villa LL, Sichero L, Rahal P, et al (2000). Molecular variants of human papillomavirus types 16 and 18 preferentially associated with cervical neoplasia. *J Gen Virol*, **81**, 2959-68.

Watts KJ, Thompson CH, Cossart YE, Rose BR (2001). Variable oncogene promoter activity of human papillomavirus type 16 cervical cancer isolates from Australia. *J Clin Microbiol*, **39**, 2009-14.

Xi LF, Koutsky LA, Galloway DA, et al (1997). Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. *J Natl Cancer Inst*, **89**, 796-802.

Yamada T, Wheeler CM, Halpern AL, et al (1995). Human papillomavirus type 16 variant lineages in United States populations characterized by nucleotide sequence analysis of the E6, L2, and L1 coding segments. *J Virol*, **69**, 7743-53.

Yamada T, Manos MM, Peto J, et al (1997). Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective. *J Virol*, **71**, 2463-72.

Zuna RE, Moore WE, Shanesmith RP et al (2009). Association of HPV16 E6 variants with diagnostic severity in cervical cytology samples of 354 women in a US population. *Int J Cancer*, **125**, 2609-13.