

# Evaluation of the genetic variability of human papillomavirus type 52

QIANG CHEN<sup>1,2</sup>, ZHAO-YUN LUO<sup>1</sup>, MIN LIN<sup>1</sup>, LIN YANG<sup>1</sup>, LI-YE YANG<sup>1</sup> and GUI-ZHI JU<sup>2</sup>

<sup>1</sup>Laboratory Medical Center, Chaozhou Central Hospital, Chaozhou, Guangdong;

<sup>2</sup>Department of Radiobiology, School of Public Health, Jilin University, Changchun, Jilin, P.R. China

Received February 28, 2012; Accepted April 30, 2012

DOI: 10.3892/ijmm.2012.1017

**Abstract.** Human papillomavirus (HPV) type 52 is one of the high-risk HPV types. Its variants can be classified as of Asian and European lineages, while little data of HPV 52 variants are available from China. In this study, the complete E6 and L1 genes were amplified and sequenced in 79 samples from eastern Guangdong, China. In total, 21 variants were identified, 18 of which could be divided into Asian lineage and the other three were phylogenetically related to European lineage. No significant difference was found between the pathogenicity of these two lineage variants. Two HPV-52 lineages variants could be classified by high resolution melting (HRM) analysis. Thus, we believe that the application of HRM analysis would be useful for facilitating pathogenic comparisons between different HPV variants. In addition, we found that 6 females were infected with two types of HPV-52 variants. To our knowledge, this is the first time that this rare phenomenon is reported worldwide.

## Introduction

Cervical cancer is the second most common gynecologic malignancy in Chinese women (1). Persistent infection of specific types of genital human papillomaviruses (HPVs), especially the oncogenic or high-risk (HR) types, has been generally recognized as the principle cause of this disease and its precursor, cervical intraepithelial neoplasia (CIN). HPV-52 is one of 13 anogenital tract infected HR-HPV types (2). According to a meta-analysis in 2003 (3), it is the seventh most common detected HR-HPV type in invasive cervical cancer (ICC) worldwide. However, in South-East Asia, HPV-52 is one of most

common prevalent HR-HPV types, and a large proportion of cervical cancers is associated with HPV-52, which is the second or third most common HPV type in ICC (4,5).

Each HPV type comprises numerous genomic variants, and an up to 2% nucleotide difference has often been identified. This difference confers each variant a biologically distinct characterization and pathogenic risks (6). HPV-52 is phylogenetically related to HPV-16 (7), whose variants could be divided into Asian-American, African and European lineages, whose carcinogenic risks have been extensively studied (8,9). Similarly, the variants of HPV-52 could be classified as Asian and European lineages (10), with a mean sequence difference of 1.8-2.0% (11). However, limited data of HPV-52 variants is available from the mainland of China (12).

The genomic characterization of HPV variants is necessary for understanding the intrinsic geographical relations and contribute to research of their pathogenicity. Traditional studies on the variability of HPV types are very laborious and costly. First, all target genes need to be sequenced following a set of polymerase chain reaction (PCR), and then the resulting sequences need to be compared with the reference sequence to identify mismatches. The high resolution melting (HRM) analysis provided scientists a new approach to detect and classify HPV variant in recent years. This kind of method combines DNA amplification by PCR and subsequent melting of the amplicons into one reaction. After data normalization, the sequence variation could be identified by the shape of melting curves (13). This new approach has been successfully used to study the variability of HPV type 16 (14).

Based on an epidemiologic screening for HPV infections in Chaozhou, eastern Guangdong province from 2009-2010, HPV-52 was found to be the most common HR-HPV type in this area. More than one third (971/2907) of HR-HPV infectors were singly or multiply infected with HPV-52. The present study explored the nucleotide variability and phylogeny of HPV-52, in samples from the HPV infections screening mentioned above. For this purpose, the E6 and L1 genes were firstly sequenced by the traditional method, then the HRM analysis was used to group identified variants. The pathogenicity of the most common variants was compared. Unexpectedly, several abnormal peaks (double peaks) were found in direct sequencing results. We speculated these specimens may contain two or more different HPV-52 variants, and the restriction enzyme analysis was performed to validate this hypothesis.

---

*Correspondence to:* Dr Li-Ye Yang, Laboratory Medical Center, Chaozhou Central Hospital, 84 Huanchengxi Road, Chaozhou, Guangdong, P.R. China  
E-mail: yangleeyee@sina.com

**Key words:** human papillomavirus type 52, genetic variability, high resolution melting, multiple variants infection

Table I. Primers for E6 and L1 genes amplification.

Gene	Target	Primer name	Sense	Antisense	Annealing temperature (°C)
E6	102-548	52E6Set1	92-113	686-713	61
		52E6Set2	26-48	550-574	57
L1	5593-7182	52L1Set1-A	5517-5537	6024-6042	57
		52L1Set1-B	5590-6007	6515-6536	57
		52L1Set1-C	6513-6533	7091-7108	57
		52L1Set1-D	6867-6887	7290-7311	58
		52L1Set2-A	5517-5537	6024-6042	57
		52L1Set2-B	5590-6007	6515-6536	57
		52L1Set2-C	6513-6533	7091-7108	57
		52L1Set2-D	6865-6887	7416-7438	59
		52L1Set3-A	5517-5537	6024-6042	57
		52L1Set3-B	5953-5972	6645-6664	58
		52L1Set3-C	6572-6591	7205-7225	60

The prototype of HPV-52 isolate TJ49-52 (GenBank ID: GQ472848) was used as the standard for nucleotide position numbering.

## Materials and methods

**Sample collection.** From December 2009 to September 2010, an epidemiologic screening for HPV infections was organized by the Chaozhou municipal government, and more than 48,000 females participated in this screening. Multiplex real-time PCR was firstly performed to detect 13 types of HR-HPV infection. Then, the HPV GenoArray test (15,16) was used to identify the specific HPV types with the same samples. HPV DNA positive women were advised to receive ThinPrep liquid-based cytology test (LCT), and histological diagnosis was performed if necessary. In this study, a woman was eligible to participate in the study if she i) participated in the HPV infection screening described above; ii) was singly infected with HPV-52; iii) received LCT and/or histological diagnosis; and iv) was willing to be a subject in the present study. The study was carried out with the approval of the Ethics Committee of Chaozhou Central Hospital, Chaozhou, China, and patient consent was obtained for the collection of cervical exfoliated cells.

In accordance with the suggestion of gynaecological practitioners, 186 HPV-52 infectors received LCT. Among them, 102 women were excluded since they were infected with multiple HPV types or did not agree to participate in the study. In total, 84 eligible cases of single HPV-52 infection were enrolled into the study.

**DNA extraction and amplification.** The cervical exfoliated cells were collected using plastic cervical swabs, and then immediately stored at 4°C. The DNA was extracted by the alkaline lysis method using DNA extraction kits (HybriBio Biotechnology Corp.). Detailed protocols for this assay have been described previously (15). The quality of extracted DNA was checked by PCR amplification of the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) gene (forward, 5'-CAT GAC CAC AGT CCA TGC CAT CAC T-3' and reverse primer, 5'-TGA GGT CCA CCA CCC TGT TGC TGT A-3').

For complete E6 and L1 gene amplification, two and three sets of primers (Table I) were designed respectively based

on the sequence of the prototype TJ49-52 (GenBank ID: GQ472848). The complete 1.8 kb genomes of the L1 gene were amplified in three or four overlapping fragments according to the primers used. The amplification protocol was as follows: initial denaturation for 5 min at 93°C, followed by 35-40 cycles of denaturation for 1 min at 93°C, annealing for 1 min at the appropriate temperature (Table I) and elongation for 10 min at 72°C. The amplified DNA fragments were identified by electrophoresis on 2% agarose gels stained with ethidium bromide, and submitted for sequencing of both strands at the Invitrogen Biotechnology Co., Ltd., Guangzhou, China.

**Sequence analysis and phylogenetic evaluations.** The mismatches were analyzed and determined using the Blast 2.0 software server (<http://www.ncbi.nlm.nih.gov/blast>). The HPV-52 prototype TJ49-52 was used as standard for comparison. Each unique sequence variation was verified by repeated amplification and opposite strand sequencing. The resulting sequences were aligned with ClustalX software, the neighbor-joining and unweighted pair group method with arithmetic average (UPGMA) trees were constructed by using the MEGA software version 5 (17,18).

**Restriction enzyme analysis.** Abnormal peaks (double peaks) were found in seven sequence traces, and their corresponding PCR products were named Ab1 to Ab7, respectively. In order to eliminate samples cross-contamination, the DNA of these specimens was re-extracted, amplified and sequenced again.

For this rare phenomenon, we speculated that the PCR products contained two kinds of amplicons. In order to validate this hypothesis, restriction enzymes *TaqI*, *XbaI*, *BspEI*, *RsaI*, *MspI* and *MlyI* (all purchased from Fermentas) were selected according to the neighboring sequence of abnormal peaks. The selection criteria were i) either one hypothetical amplicon or another was cleavable; and ii) there was only one recognition site of restriction enzyme on the selected amplicon. The PCR products were digested with specific enzymes according to the recommended protocols for digestion of PCR products

Table II. Details for restriction enzyme analysis.

PCR product	Primers	Amplicon length	Restriction enzyme	Recognition site	Digestion	Inactivation	Fragment length
Ab1	52E6Set2	549	<i>RsaI</i>	5'-GT↓AC-3'	37°C for 24 h	80°C for 20 min	198, 351
Ab2-5 <sup>a</sup>	52L1Set1-A	526	<i>TaqI</i>	5'-T↓CGA-3'	65°C for 3 h	EDTA solution <sup>b</sup>	252, 274
			<i>XbaI</i>	5'-T↓CTAGA-3'	37°C for 24 h	65°C for 20 min	254, 272
Ab6	52L1Set1-A	526	<i>BspEI</i>	5'-T↓CCGGA-3'	55°C for 3 h	80°C for 20 min	108, 418
Ab7	52L1Set3-B	712	<i>MspI</i>	5'-C↓CGG-3'	37°C for 24 h	80°C for 20 min	518, 194
			<i>MlyI</i>	3'-CTCAG(n)5'±-5'	37°C for 8 h	65°C for 20 min	532, 180

<sup>a</sup>For amplicons Ab2 to Ab5, the restriction enzymes used for digestion are identical, since the same double peaks are identified at the same site in their sequence traces. <sup>b</sup>Adding 0.5 M EDTA, pH 8.0, to achieve a 20 mM final concentration. Arrows (↓) indicate the cleavage site.

Table III. The GenBank IDs for the isolates reported in this study.

Isolates	GenBank ID for E6 gene	GenBank ID for L1 gene	Lineage
CZ52A105	JN874437	JN874416	Asian
CZ52A207	JN874438	JN874417	Asian
CZ52A255	JN874439	JN874418	Asian
CZ52A620	JN874441	JN874420	Asian
CZ52A1404	JN874442	JN874421	Asian
CZ52A1230	JN874443	JN874422	Asian
CZ52A375	JN874444	JN874423	Asian
CZ52A1569	JN874446	JN874425	Asian
CZ52A264	JN874447	JN874426	Asian
CZ52A247	JN874448	JN874427	Asian
CZ52A277	JN874449	JN874428	Asian
CZ52A336	JN874450	JN874429	Asian
CZ52A1023	JN874451	JN874430	Asian
CZ52A228	JN874453	JN874432	Asian
CZ52A168	JN874454	JN874433	Asian
CZ52D416	JN874455	JN874434	Asian
CZ52D417	JN874456	JN874435	Asian
CZ52A921	JN874457	JN874436	Asian
CZ52E928	JN874440	JN874419	European
CZ52E136	JN874445	JN874424	European
CZ52E429	JN874452	JN874431	European

directly after amplification. The 50  $\mu$ l digestion mixture included 15  $\mu$ l PCR reaction mixture, 29  $\mu$ l nuclease-free water, 3  $\mu$ l 10X buffer and 3  $\mu$ l (10 U/ $\mu$ l) restriction enzyme. The detailed protocols for digestion are listed in Table II. The resulting fragments were run on 2% agarose gels stained with ethidium bromide, and the uncleaved fragments were purified and submitted for sequencing.

**HRM analysis.** All of the specimens which could be successfully amplified with PCR, were subjected to HRM analysis. The primers (forward, 5'-TGT ATT ATG TGC CTA CGC TTT T-3' and reverse, 5'-GGC GTT TGA CAA ATT ATA CAT C-3') used for HRM analysis were designed based on the mismatches identified on the E6 gene. PCR reactions were performed on the LightCycler® 480 II. The instrument (Roche Diagnostics) was equipped with the software LightCycler® 480 Gene Scanning

Software Version 1.5 (Roche Diagnostics). Approximately 50 ng of DNA was amplified in a total volume of 20  $\mu$ l containing 0.2  $\mu$ M of each primer, 1X PCR buffer, 0.2 mM of each dNTP, 0.3 unit TaqDNA polymerase (Dongsheng Biotech Co., Ltd. Guangzhou, China) and 1  $\mu$ l LC Green plus (Idaho Technology). The reaction conditions were 95°C for 3 min, followed by 35 cycles at 98°C for 20 sec, 60°C for 20 sec, and 72°C for 20 sec. The melting program included three steps: denaturation at 95°C for 1 min, renaturation at 40°C for 1 min and then melting in a continuous fluorescent reading from 60-95°C at 20 acquisitions/°C.

**Liquid-based cytology test (LCT) and pathological diagnosis.** LCT was performed in the Chaozhou Central Hospital. The detailed protocols were previously described (15). The results were evaluated using the Bethesda system (19). The evaluation

Table IV. Mismatches on the E6 gene.

Isolate	Nucleotide position										
	90	136	168	228	249	255	277	278	336	366	429
CZ52A105	T	C	C	T	T	G	A	G	T	C	A
CZ52A1569	C	C	C	T	T	G	A	G	T	C	A
CZ52A336	T	C	C	T	T	G	A	G	C	C	A
CZ52A228	T	C	C	C	T	G	A	G	T	C	A
CZ52A255	T	C	C	T	T	A	A	G	T	C	A
CZ52A168	T	C	T	T	T	G	A	G	T	C	A
CZ52A277	T	C	C	T	T	G	C	G	T	A <sup>a</sup>	A
CZ52E928	T	C	C	T	G	G	A	A <sup>b</sup>	T	C	A
CZ52E429	T	C	C	T	T	G	A	A <sup>b</sup>	T	C	G
CZ52E136	T	T	C	T	T	G	A	A <sup>b</sup>	T	C	G

The E6 gene sequence of isolate CZ52A105 (GenBank ID, JN874437) is used as the standard for comparison and nucleotide position numbering. The E6 gene sequences of isolates CZ52A207, CZ52A620, CZ52A1404, CZ52A1230, CZ52A375, CZ52A247, CZ52A1023, CZ52D416 and CZ52A921 completely match those of isolate CZ52A105, and are not listed. The E6 gene sequences of isolates CZ52A264 and CZ52D417 are identical with CZ52A255, and are not listed. <sup>a</sup>Non-synonymous mutation, leading to the change of amino acid R122K. <sup>b</sup>Non-synonymous mutation, leading to the change of amino acid N93K. Other mismatches are all synonymous mutations.

system included: i) negative (A0), ii) atypical squamous cells (ASC), iii) low-grade squamous intraepithelial lesion (LSIL), iv) high-grade squamous intraepithelial lesion (HSIL), and v) squamous cell carcinoma (SCC). The LSIL, HSIL and SCC cases further received biopsies and the samples were processed and diagnosed in the Department of Pathology, Chaozhou Central Hospital.

**Statistical analysis.** For pathogenic risks assessment, binary and multinomial logistic regression analysis was used. Firstly, the pathogenic risks were compared (binary logistic regression) between two main lineages variants according to the phylogenetic trees constructed. Then, the pathogenic risks of the four most common variants were compared (multinomial logistic regression). All data were analyzed using SPSS software version 16. P-values were two-sided, and statistical significance was accepted if the P-value was  $\leq 0.05$ .

## Results and Discussion

**Phylogenetic and geographical relatedness of HPV52 variants.** Of the total 84 specimens, five specimens were excluded since they were negative for GAPDH amplification, and the complete E6 and L1 genes were amplified and sequenced successfully in 79 cases. Among them, six cases were not suitable for genetic variability evaluation because they were infected with two kinds of HPV-52 variants. Thus, the HPV-52 genetic variability was evaluated in 73 HPV-52 singly infected females (median age 44.5 years, range 35.5-59.4 years).

At the nucleotide level, a total of 21 HPV-52 variants were identified when E6 and L1 sequences were combined and the prototype TJ49-52 was used as the standard for comparison. The sequences of these 21 variants have been submitted to the GenBank. The GenBank IDs are JN874437-JN874457 for the E6 gene and JN874416-JN874436 for the L1 gene (Table III). The maximum nucleotide diversity was 1.1% (22/2037) observed between isolates CZ52A375 and CZ52E429.

There were 10 and 15 kinds of variability found on E6 and L1 gene, respectively. Across the E6 gene 2.5% nucleotide sites (11/447) and 1.3% encoded amino acids (2/149) were variable (Table IV), while 1.9% nucleotide sites (31/1590) and 1.1% encoded amino acids (6/530) were variable across the L1 gene (Table V).

Great similarity was found between the phylogenetic trees constructed based on the E6 or L1 gene. Partial and complete HPV-52 sequence reports in different regions of the world formed evolutionary trees with two main branches driven by variants with high prevalence in Asia (Fig. 1, branch A) or Europe (Fig. 1, branch E), which was coincident with the previous study (10). The details of the sequences used for phylogenetic tree construction are presented in Table VI.

The isolate CZ52A105 represented the most common variant on branch A with 45.2% (33/73) of the cases infected with this kind of variant (Table VII). Following was isolate CZ52A255 (8/73) and CZ52A207 (8/73). There was one synonymous mutation compared with CZ52A105 on E6 for the former and on L1 for the latter. The fourth most common variant were isolates CZ52A228 (2/73) and CZ52A264 (2/73). There was one synonymous mutation compared with CZ52A105 on E6 for the former, and one synonymous mutation on E6 and L1 respectively for the latter. The sequences of other variants located on branch A were all unique.

The isolate CZ52A105 located at the bifurcation of branch A, 83.6% (61/73) samples were identical with CZ52A105 at amino acid level. Moreover, its E6 and L1 sequences were completely matched with isolate IN141070 (11/20) reported in Chiang Mai (N 18.56, E 72.49). The E6 sequence was also identical with isolates b00422 and b00433 reported in Guangzhou (N 23.70, E 113.15) and isolate HK1243 and HK2571 (17) reported in HongKong (N 22.65, E 113.86). The latitude of Chaozhou (N 23.40, E 116.38) and of these 3 cities all ranged from N 18 to N 24. Taken all above mentioned into consideration, we suggest that the isolate CZ52A105 represents the most common and ancient HPV-52 variant in South-East Asia.

Table V. Mismatches on the L1 gene.

Isolate	Nucleotide position																															
	14	207	247	264	345	375	408	444	519	546	620	621	654	879	921	975	999	1023	1134	1137	1192	1200	1230	1246	1260	1353	1377	1404	1488	1548	1569	
CZ52A105	T	G	A	A	A	A	C	G	G	A	A	T	A	C	A	T	A	G	G	G	C	C	G	G	T	A	A	A	A	G	G	
CZ52A207	T	A	A	A	A	A	C	G	G	A	A	T	A	C	A	T	A	G	G	G	C	C	G	G	T	A	A	A	A	G	G	
CZ52A620	T	G	A	A	A	A	C	G	G	A	<u>C<sup>a</sup></u>	T	A	C	A	T	A	G	G	G	C	C	G	G	T	A	A	A	A	G	G	
CZ52A1404	T	G	A	A	A	A	C	G	G	A	A	T	A	C	A	T	A	G	G	G	C	C	G	G	T	A	A	<u>C<sup>b</sup></u>	A	G	G	
CZ52A1230	T	G	A	A	A	A	C	G	G	A	A	T	A	C	A	T	A	G	G	G	C	C	A	G	T	A	A	A	A	G	G	
CZ52A1569	T	G	A	A	A	A	C	G	G	A	A	T	A	C	A	T	A	G	G	G	C	C	G	G	T	A	A	A	A	G	A	
CZ52A264	T	G	A	<b>G</b>	A	A	C	G	G	A	A	T	A	C	A	T	A	G	G	G	C	C	G	G	T	A	A	A	A	G	G	
CZ52A1023	T	G	A	A	A	A	C	G	G	A	A	T	A	C	A	T	A	A	G	G	C	C	G	G	T	A	A	A	A	G	G	
CZ52A277	T	G	A	A	A	A	C	G	G	A	A	T	A	C	A	T	A	G	G	G	C	C	G	<b>A<sup>c</sup></b>	T	A	<b>G</b>	A	A	G	G	
CZ52A921	T	<b>A</b>	A	A	A	A	C	G	G	A	A	T	A	C	<b>G</b>	T	A	G	G	G	C	C	G	G	T	A	A	A	A	G	G	
CZ52A247	T	G	<b>G<sup>d</sup></b>	A	A	A	C	A	G	A	A	T	A	C	<b>G</b>	A	A	G	G	G	C	C	G	G	T	A	A	A	A	G	G	
CZ52A375	T	G	A	<b>G</b>	A	<b>G</b>	C	G	G	A	<u>C<sup>a</sup></u>	C	A	C	A	T	<b>G</b>	G	G	<b>A<sup>e</sup></b>	C	C	G	G	T	A	A	A	A	G	G	
CZ52D416	T	G	A	A	A	A	C	G	A	<b>G</b>	A	T	A	T	A	T	A	G	G	G	C	C	G	G	T	A	A	A	A	G	G	
CZ52E928	T	<b>A</b>	A	A	A	A	A	T	G	<b>G</b>	A	T	<b>G</b>	C	A	T	A	G	G	T	<b>G</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>G</b>	<b>G</b>	<b>G</b>	
CZ52E429	<b>C<sup>f</sup></b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>T</b>	<b>G</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>T</b>	<b>A</b>	<b>T</b>	<b>A</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>T</b>	<b>G</b>	<b>T</b>	<b>A</b>	<b>G</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>G</b>	<b>A</b>	<b>G</b>	

The L1 gene sequence of isolate CZ52A105 (GenBank ID: JN874416) was used as the standard for comparison and nucleotide position numbering. The L1 gene sequences of isolates CZ52A255, CZ52A168, CZ52A228 and CZ52A336 completely matched with isolate CZ52A105, and are not listed. The L1 gene sequence of isolate CZ52D417 is identical with CZ52D416, and is not listed. The L1 gene sequence of isolate CZ52E136 is identical with CZ52E429, and is not listed. <sup>a</sup>Non-synonymous mutation, leading to the change of amino acid N207T. <sup>b</sup>Non-synonymous mutation, leading to the change of amino acid E468D. <sup>c</sup>Non-synonymous mutation, leading to the change of amino acid D416N. <sup>d</sup>Non-synonymous mutation, leading to the change of amino acid S83G. <sup>e</sup>Non-synonymous mutation, leading to the change of amino acid E398K. <sup>f</sup>Non-synonymous mutation, leading to the change of amino acid L5S. Other mismatches are all synonymous mutations.

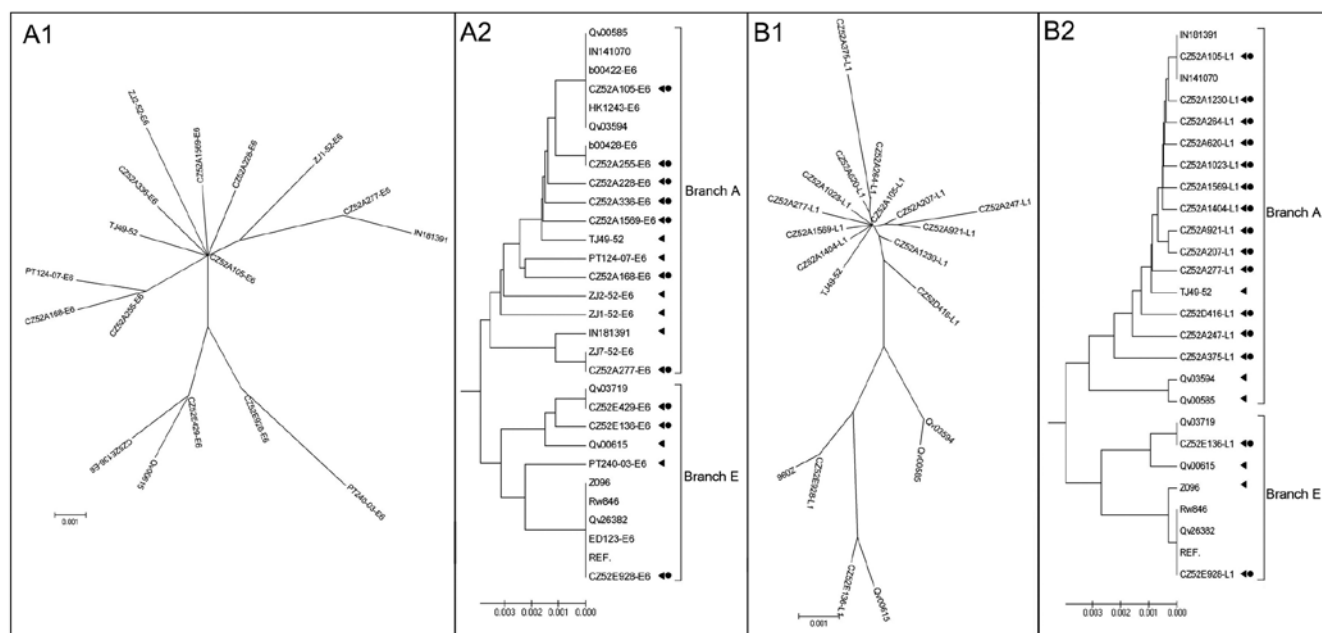


Figure 1. The intratype diversity of HPV-52. The phylogenetic trees represent the relationship between HPV-52 variants based on (A1 and A2) E6 or (B1 and B2) L1 gene. (A1 and B1) are based on the UPGMA, (A2 and B2) are based on the neighbor-joining algorithm. The UPGMA tree represents all isolates. The variants chosen to represent a particular variant in the neighbor-joining tree are indicated by black triangles. The variants reported in this study are signaled with black round. The GenBank IDs of the isolates reported in this study are shown in Table III, the details for the other sequence used for phylogenetic tree construction is shown in Table VI.

Table VI. Details of the sequences used for phylogenetic tree construction.

Isolate	Sequence type	GenBank ID	City/Country	Lineage
b00422	E6 genes, complete cds	EU924143	Guangzhou, China	Asian
b00428	E6 genes, complete cds	EU924144	Guangzhou, China	Asian
HK1243	E6 genes, complete cds	DQ057293	Hongkong, China	Asian
ZJ1-52	E6 genes, complete cds	HM004116	Hangzhou, China	Asian
ZJ2-52	E6 genes, complete cds	HM004115	Hangzhou, China	Asian
ZJ7-52	E6 genes, complete cds	HM004117	Hangzhou, China	Asian
PT124-07	E6 genes, complete cds	FJ002423	Portugal	Asian
TJ49-52	Complete genome	GQ472848	Tianjin, China	Asian
IN141070	Complete genome	HQ537743	Chiang Mai, Thailand	Asian
IN181391	Complete genome	HQ537742	Chiang Mai, Thailand	Asian
Qv00585	Complete genome	HQ537741	Guanacaste, Costa Rica	Asian
Qv03594	Complete genome	HQ537740	Guanacaste, Costa Rica	Asian
ED123	E6 genes, complete cds	DQ057290	Scotland	European
PT240-03	E6 genes, complete cds	EF566934	Portugal	European
REF.	Complete genome	X74481	Germany	European
Rw846	Complete genome	HQ537735	Rwandan	European
Z096	Complete genome	HQ537736	Zambia	European
Qv03719	Complete genome	HQ537745	Guanacaste, Costa Rica	European
Qv26382	Complete genome	HQ537732	Guanacaste, Costa Rica	European
Qv00615	Complete genome	HQ537746	Guanacaste, Costa Rica	European

CDS, sequence coding for amino acids in protein.

Nearly one tenth (7/73) of the cases were phylogenetically related to European lineages, and three variants were identified. The isolate CZ52E429 (4 cases) represented the most frequent variant, followed by isolate CZ52E928 (2 cases). Compared with CZ52E429, one non-synonymous mutant was observed

on L1, two and seven synonymous mutants were identified on E6 and L1 respectively. The other isolate located on branch E was CZ52E136 (1 case), there was only one synonymous mutant found on E6 if the sequence of isolate CZ52E429 was used as a reference. The isolate CZ52E928 located at the bifur-

Table VII. LCT results for the most frequent variants.

Isolate (lineage <sup>a</sup> )	Cases, n	LCT results, n				OR (95% CI) <sup>b</sup>
		ASC	LSIL	HSIL and SCC	LCT <sup>+</sup> (%)	
CZ52A105 (A)	33	10	3	1	14 (42.4)	Reference
CZ52A207 (A)	8	4	0	0	4 (50.2)	1.36 (0.29-6.38)
CZ52A255 (A)	8	5	0	0	5 (62.5)	2.26 (0.46-11.08)
CZ52E429 (E)	4	3	0	0	3 (75.0)	4.07 (0.38-43.39)
Total Asian variants	66	23	8	2	33 (50.0)	Reference
Total European variants	7	4	0	0	4 (57.1)	1.33 (0.28-6.43)

<sup>a</sup>A, represents Asian lineage; E, represents European lineage. <sup>b</sup>OR was obtained using binary (between two main lineages variants) and multinomial (between the 4 most common variants) logistic regression analysis model. OR, odds ratio; 95% CI, 95% confidence intervals.

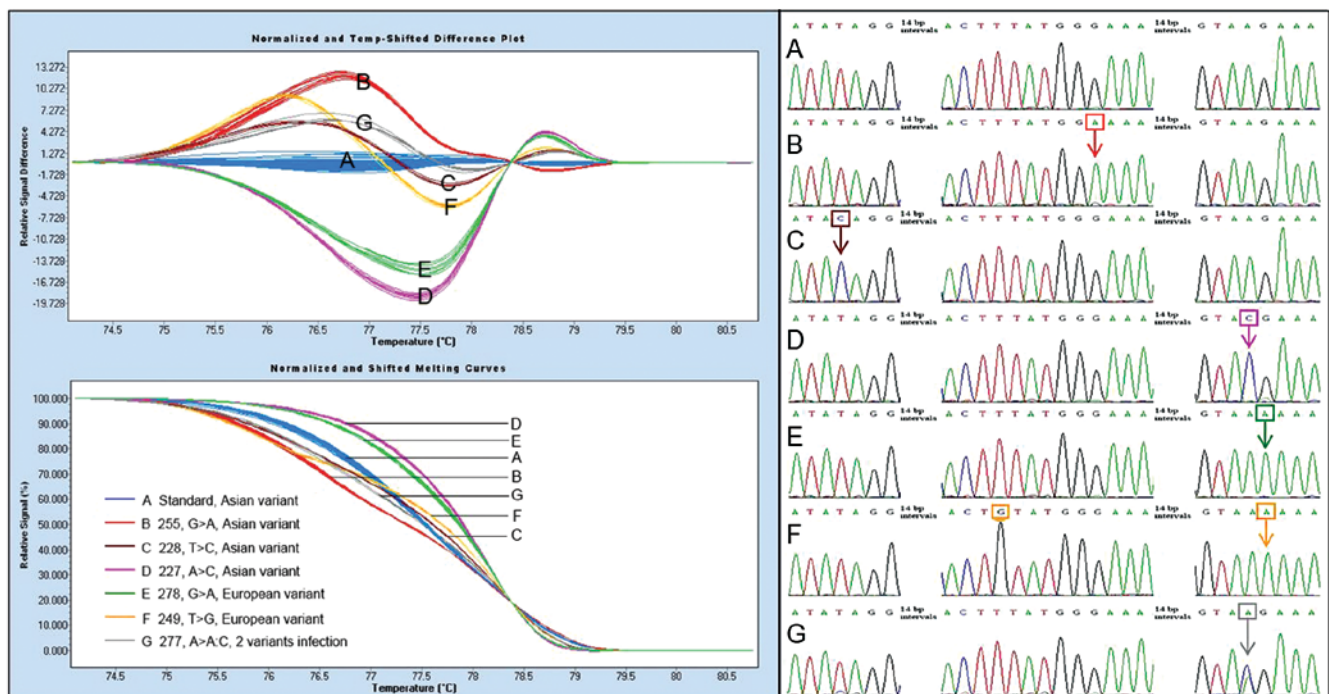


Figure 2. HPV-52 variant group identified by HRM analysis. HRM analysis used to group 79 specimens (including 6 cases with 2 variants infection). Several repeats (x n) were used for the less frequent variants. Group A, n=54x1; Group B, n=14x1; Group C, n=2x3; Group D, n=1x6; Group E, n=5x2; Group F, n=2x3; Group G, n=1x6.

cation of branch E, it may represent the most widely spread variant. The same E6 and L1 sequences were also found in Europe (Germany) (21), Africa (Rwanda) and North America (Costa Rica) (11,22,23). We speculate that this variant may be derived from the worldwide migration of the European host.

**Variants pathogenicity assessment and HRM analysis application.** Despite phylogenetic relatedness, a geographic-related pathogenicity difference has been found in HPV-16 and HPV-18 variants. The carcinogenic risk for Asian-American or African HPV16 variants is greater than that of European variants (8), and non-European HPV-18 variants are more common in cancer tissues than European variants (9).

A preliminary pathogenicity comparison between HPV-52 variants was performed in this study. The LCT results of the most frequent variants are shown in Table VII. A total of 73

specimens were firstly divided into 2 groups, one was made up of the Asian lineage related specimens (66 cases), and the other was composed of European lineage related specimens (7 cases). The pathogenic risk was assessed by binary logistic regression analysis. The median age of the two groups was 45.4 and 42.2-year-old, no significant difference was found between the pathogenic risk of these two lineages variants ( $P=0.72$ ). Then, the pathogenic risks of the four most common variants were evaluated. There was no statistically significant difference between the groups ( $P=0.51$ ). However, there were some limitations for this conclusion, because there was a great unbalance between the case numbers of the variants. In order to validate this conclusion, it is necessary to perform this experiment in a larger sample.

The HRM analysis provided us a more economic and rapid method to deal with large scale screening HPV-52 variants. As a



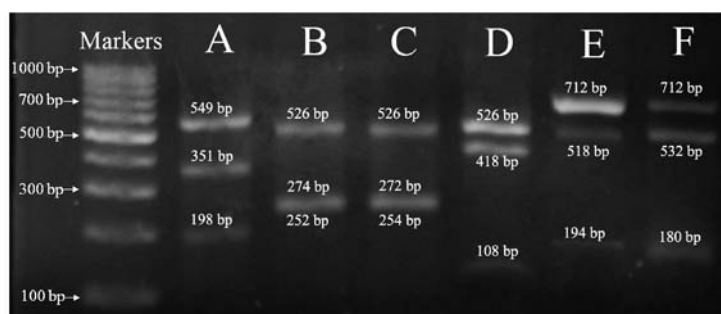


Figure 3. Products of restriction enzyme digestion. (Lane A) Ab1 digestion with *RsaI*; (lane B) Ab2 digestion with *TaqI*; (lane C) Ab2 digestion with *XbaI*; (lane D) Ab6 digestion with *BspEI*; (lane E) Ab7 digestion with *MspI*; (lane F) Ab7 digestion with *MlyI*. The digestion products of Ab3, Ab4 and Ab5 are identical with Ab2.

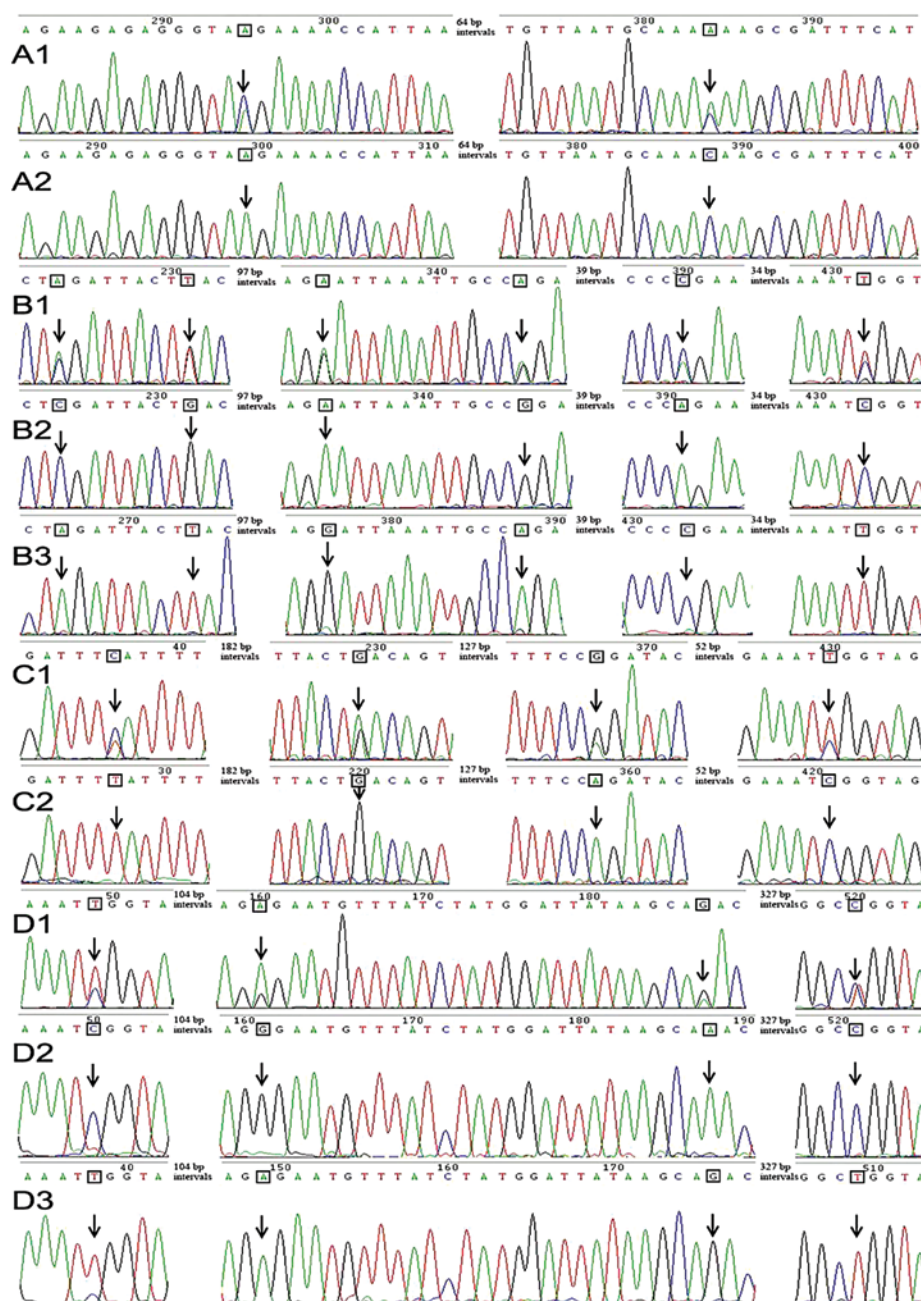


Figure 4. Verification of multiple variants infection. (A1, B1, C1 and D1) indicate the double peaks on the direct sequencing results of Ab1, Ab2, Ab6 and Ab7, respectively; the others are sequence traces of the uncleavable products after digestion with specific restriction enzyme. (A2) Ab1 digestion with *RsaI*; (B2) Ab2 digestion with *XbaI*; (B3) Ab2 digestion with *TaqI*; (C2) Ab6 digestion with *BspEI*; (D2) Ab7 digestion with *MlyI*; (D3) Ab7 digestion with *MspI*. The sequence details of Ab3, Ab4 and Ab5 are identical with Ab2.



preliminary attempt, the E6 gene was used to group identified HPV-52 variants by HRM analysis. The choice of E6 gene for the analysis of the variability by HRM was based on the fact that it was the most common gene examined by sequencing worldwide, and because it has an optimal variable region (6,14).

A shorter amplicon was usually required for HRM analysis because as amplicon size increase, it becomes increasingly difficult to identify mismatches (13). For this reason, the amplicon (nucleotide 187-326, 140 bp) used for HRM analysis was only one third of the complete E6 gene. This amplicon involved five mismatches found on the E6 gene in our study, and these mismatches led to six kinds of variants. One variant involved two mismatches, the other five contained only one mismatch. It is worth pointing out that the consensus mismatch between the Asian and European lineages was located at nucleotide 278 (G>A) (Table IV), and it was also located in the amplicon.

After all the data were normalized, a total of 79 specimens (including six multiple variants infection cases) were divided into six groups (Fig. 2), and all of six variants on the amplicon could be divided into distinct groups. Four groups (Fig. 2A-D) were composed of Asian variants, and the other two (Fig. 2E and F) were made of European variants. In this sense, Asian and European variants could be distinguished by HRM analysis with a complete accuracy. Therefore, the application of HRM would facilitate the HPV-52 variants pathogenic comparison between Asian and European lineages. In addition, one case with two HPV-52 variants infection (Ab1) could also be identified by the shape of the resulting melting curve (Fig. 2G).

**Multiple variant infections.** In this study, some double peaks were found in 7 direct sequencing results, and 6 specimens were involved in these events. The Ab6 and Ab7 were consecutive fragments derived from one specimen, and the other 5 came from different specimens. Interestingly, the same double peaks were found in four of them (Ab2-Ab5) at the same nucleotide position.

After digestion with specific restriction enzymes, three electrophoresis bands could be found, including two new generated bands and one uncleavable band (Fig. 3), which was purified and submitted for sequencing. The double peaks disappeared in the sequence trace of the uncleavable fragments, which were replaced by a single base at the corresponding positions (Fig. 3). Using Ab2 (Fig. 3C) as an example, six double peaks were observed in its direct sequencing results (Fig. 4B2). They were composed of peaks, C:A, G:T, A:G, G:A, A:C and C:T in turn. After digestion with restriction enzyme *Xba*I, these double peaks were replaced with single base C, G, A, G, A and C respectively in the sequence trace of the uncleavable fragment. On the contrary, if this PCR product was digested with restriction enzyme *Taq*I, these double peaks were replaced with single base A, T, G, A, C and T, respectively. Based on this result, we believed that the most possible reason for this event was that the specimens contained two HPV-52 variants, in other words, these women infected two kinds of HPV-52 variants. We believe that this phenomenon may be similar to the multiple type infections. Once a woman is singly infected with HPV-52, she has an almost equal opportunity of further infection with other HPV types or infection with HPV-52 again,

the former events leading to multiple infections, and the latter possibly leading to two kinds of HPV-52 variants infection.

To our knowledge, this rare phenomenon has not so far been reported worldwide. However, since it was observed in 7.6% (6/79) specimens in this study, why was it not observed in the previous studies? We think that there are three possible reasons. First, this rare phenomenon is preferentially observed in an area of high HPV-52 prevalence. Second, the primers used for amplification must have been suitable for two kinds of variants. Last but not least, the viral loads of two different variants should be nearly equal in the same sample. However, the lower one might be ignored in the sequencing trace if there was a great difference between the viral loads of the two variants. As described above, HPV-52 was the most common HPV type in eastern Guangdong, and its high prevalence may increase the opportunity of multiple variants infection.

## Acknowledgements

This study was supported by grants by the Medical Research Funds of Guangdong Province (A2011760, B2008179). Reagent discounts and reagent gifts were from the HybriBio Biotechnology Limited Corp. We offer special recognition for the excellent work of the study staff in sample collection. We sincerely thank SY Zheng and JB Liu (Department of Pathology, Chaozhou Central Hospital) for technical assistance with pathology.

## References

1. Chen HC, Schiffman M, Lin CY, Pan MH, You SL, Chuang LC, Hsieh CY, Liaw KL, Hsing AW, Chen CJ; CBCSP-HPV Study Group: Persistence of type-specific human papillomavirus infection and increased long-term risk of cervical cancer. *J Natl Cancer Inst* 103: 1387-1396, 2011.
2. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, Snijders PJ, Meijer CJ; International Agency for Research on Cancer Multicenter Cervical Cancer Study Group: Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 348: 518-527, 2003.
3. Clifford GM, Smith JS, Plummer M, Muñoz N and Franceschi S: Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *Br J Cancer* 88: 63-73, 2003.
4. Ho CM, Chien TY, Huang SH, Lee BH and Chang SF: Integrated human papillomavirus types 52 and 58 are infrequently found in cervical cancer, and high viral loads predict risk of cervical cancer. *Gynecol Oncol* 102: 54-60, 2006.
5. Takehara K, Toda T, Nishimura T, Sakane J, Kawakami Y, Mizunoe T, Nishiwaki M and Taniyama K: Human papillomavirus types 52 and 58 are prevalent in uterine cervical squamous lesions from Japanese women. *Patholog Res Int* 2011: 246936, 2011.
6. Bernard HU, Calleja-Macias IE and Dunn ST: Genome variation of human papillomavirus types: phylogenetic and medical implications. *Int J Cancer* 118: 1071-1076, 2006.
7. Schiffman M, Herrero R, Desalle R, Hildesheim A, Wacholder S, Rodriguez AC, Bratti MC, Sherman ME, Morales J, Guillen D, et al: The carcinogenicity of human papillomavirus types reflects viral evolution. *Virology* 337: 76-84, 2005.
8. Hildesheim A, Schiffman M, Bromley C, Wacholder S, Herrero R, Rodriguez A, Bratti MC, Sherman ME, Scarpidis U, Lin QQ, et al: Human papillomavirus type 16 variants and risk of cervical cancer. *J Natl Cancer Inst* 93: 315-318, 2001.
9. Sichero L, Ferreira S, Trottier H, Duarte-Franco E, Ferenczy A, Franco EL and Villa LL: High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18. *Int J Cancer* 120: 1763-1768, 2007.
10. Raiol T, Wyant PS, de Amorim RM, Cerqueira DM, Milanezi NG, Brígido Mde M, Sichero L and Martins CR: Genetic variability and phylogeny of the high-risk HPV-31, -33, -35, -52, and -58 in central Brazil. *J Med Virol* 81: 685-692, 2009.

11. Chen Z, Schiffman M, Herrero R, Desalle R, Anastos K, Segondy M, Sahasrabudhe VV, Gravitt PE, Hsing AW and Burk RD: Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. *PLoS One* 6: e20183, 2011.
12. Wu XL, Zhang CT, Zhu XK and Wang YC: Detection of HPV types and neutralizing antibodies in women with genital warts in Tianjin City, China. *Viol Sin* 25: 8-17, 2010.
13. Wittwer CT, Reed GH, Gundry CN, Vandersteen JG and Pryor RJ: High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin Chem* 49: 853-860, 2003.
14. Sabol I, Cretnik M, Hadzisejdić I, Si-Mohamed A, Matovina M, Grahovac B, Levanat S and Grce M: A new approach for the evaluation of the human papillomavirus type 16 variability with high resolution melting analysis. *J Virol Methods* 162: 142-147, 2009.
15. Lin M, Yang LY, Li LJ, Wu JR, Peng YP and Luo ZY: Genital human papillomavirus screening by gene chip in Chinese women of Guangdong province. *Aust NZ J Obstet Gynaecol* 48: 189-194, 2008.
16. Liu SS, Leung RC, Chan KK, Cheung AN and Ngan HY: Evaluation of a newly developed GenoArray human papillomavirus (HPV) genotyping assay and comparison with the Roche Linear Array HPV genotyping assay. *J Clin Microbiol* 48: 758-764, 2010.
17. Calleja-Macias IE, Villa LL, Prado JC, Kalantari M, Allan B, Williamson AL, Chung LP, Collins RJ, Zuna RE, Dunn ST, *et al*: 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, 2005.
18. Tamura K, Peterson D, Peterson N, Stecher G, Nei M and Kumar S: MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739, 2011.
19. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, Raab S, Sherman M, Wilbur D, Wright T Jr, Young N; Forum Group Members; Bethesda 2001 Workshop: The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* 287: 2114-2119, 2002.
20. Wongworapat K, Keawvichit R, Sirirojn B, Dokuta S, Ruangyuttikarn C, Sriplienchan S, Sontirat A, Kla KT, Gravitt PE and Celentano DD: Detection of human papillomavirus from self-collected vaginal samples of women in Chiang Mai, Thailand. *Sex Transm Dis* 35: 172-173, 2008.
21. Delius H and Hofmann B: Primer-directed sequencing of human papillomavirus types. *Curr Top Microbiol Immunol* 186: 13-31, 1994.
22. Herrero R, Castle PE, Schiffman M, Bratti MC, Hildesheim A, Morales J, Alfaro M, Sherman ME, Wacholder S, Chen S, *et al*: Epidemiologic profile of type-specific human papillomavirus infection and cervical neoplasia in Guanacaste, Costa Rica. *J Infect Dis* 191: 1796-1807, 2005.
23. Singh DK, Anastos K, Hoover DR, Burk RD, Shi Q, Ngendahayo L, Mutimura E, Cajigas A, Bigirimani V, Cai X, *et al*: Human papillomavirus infection and cervical cytology in HIV-infected and HIV-uninfected Rwandan women. *J Infect Dis* 199: 1851-1861, 2009.