Genetic Variability and Phylogeny of Human Papillomavirus Type 16 Based On E6, E7 and L1 Genes in Central China

Shuizhong Han
Northwest Agriculture University: Northwest Agriculture and Forestry University
Xiaochuan Wang
989th Hospital of PLA
Xiaojing Wang
989th Hospital of PLA
Shuaijun Wang
989th Hospital of PLA
Li Ma (sunshinejiayuan@163.com)
989th Hospital of PLA https://orcid.org/0000-0002-5444-5262

Research Article

Keywords: Human papillomavirus16, sublineage, Genetic variability, phylogeny

Posted Date: November 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1090848/v1

Abstract

In the current study, a total of 74 single-infected HPV16 samples from females attending the gynecological outpatient clinic in four cities of Henan province were collected and applied to the L1, E6 and E7 sequencing. Variations of the HPV16 L1, E6 and E7 genes were characterized by comparison with reference sequence and the secondary structure analysis were conducted. Phylogenetic trees based on the L1 and E6-E7 sequences were constructed separately. B-cell epitopes of the HPV16 E6 and E7 proteins were predicted further. A total of thirty-seven novel variations, including twenty L1 genes and seventeen E6-E7 genes were identified. Compared with the reference sequence, twenty-eight variations (1.8%, 28/1596) were identified in L1 gene sequences and 10/28 (35.7%) were non-synonymous mutations. For E6-E7 sequences, twenty-five novel gene changes (including 16 mutations (3.4%, 16/477) in E6 gene, 9 mutations (3.0%, 9/297) in E7 gene) were found, 18/25 (72.0%) were non-synonymous and 10/28 (35.7%) were non-synonymous mutations. Phylogenetic analysis showed that 56.8% (42/74) of the samples were A1 sublineages, 37.8% (28/74) were A4, 4.1% (3/74) were A3 and 1.4% (1/74) was A2 sublineages. On the prediction of B-cell epitopes, seven potent epitopes for E6 and four for E7 were identified. Amino mutations, including L90V, R62K, R142Q and F76L in E6, S63F and N29S/H in E7 changed the score. HPV16 variants prevalent in the central China belong to European A1 sublineages. Sequences of HPV16 L1, E6 and E7 in this study may provide assistant for the improvement of HPV vaccines.

Introduction

Cervical cancer is the fourth most commonly diagnosed and leading cause of cancer death among females worldwide, with an estimated 570,000 cases and 311,000 deaths in 2018 [1]. Around 85% of women diagnosed and 87% of women who died from cervical cancer live in the developing countries [2]. Human papillomavirus (HPVs) were found in the cervical carcinoma tissues of most patients and the oncogenic HPVs are regarded as the major cause of cervical cancer. In China, it was reported that there are estimated 110,650 new cancer cases and 36,714 cancer deaths are attributable to HPVs infection in 2015, of which cervical cancer accounted for 85.6% and 78.1% [3].

HPVs are small non-enveloped double-stranded DNA viruses that belong to the genus *Alpha-Papillomaviridae* family [4]. The HPVs genomes are about 7.2kb to 8.0kb and contain eight open reading frames (ORFs), including: the presumptive early (E1-E2, E4-E7), late (L1 and L2) and Long Control Region (LCR) [5–7]. The continued expression of the E6 and E7 genes is related to induce cellular immortalization, transformation, and carcinogenesis [6]. The L1 protein is the primary composition of HPVs and self-assembles into virus like particles (VLPs) [8]. The first-generation commercial HPV vaccines are based on the recombinant expression of L1 protein in system, such as the bivalent and quadrivalent HPV vaccine [9–10]. Human immunized with commercial HPV vaccines can acquire robust immunity against the homology genotype [8]. The E6 and E7 proteins could be candidate for the development of therapeutic vaccines [11].

More than 200 different HPVs genotypes have been characterized according to the greater than 10% difference within the L1 gene sequence [4, 5, 12]. Based on their association with cervical cancer, HPV genotypes are classified into high-risk HPV (HR-HPV, including HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59) and low-risk HPV (LR-HPV, including HPV6, 11, 40, 42, 43 and 44) [12]. Individual HPV genotypes are referred to as variants or subtypes when less than 10% difference is within the L1 gene sequence [5]. The HPVs variants are grouped into distinct lineages and sublineages based on nucleotides alignments and phylogenetic analyses [5]. Usually, the L1 nucleotides difference among HPV lineages and sublineages is 1.0%-10.0% and 0.5%-1.0%, respectively [5].

HPV16 is detected in 70% of cases of invasive cervical worldwide [13]. In China, HPV16 is the predominant type in most areas [14–19]. HPV16 has been classified into four major lineages and nine sublineages, including: (1) A, includes A1-3(European [E]) and A4(Asian [As]) sublineages;(2) B, includes B1 (African-1 [Afr1a]) and B2 (African-1 [Afr1b]); (3) C (African-2 [Afr2a]); and (4) D, includes D1 (North American 1 [NA]), D2 (Asian-American 2 [AA]) and D3 (Asian-American 1 [AA]) [5, 20]. The lineages and sublineages of HPV16 sequences have geography characteristic [21–23]. It was reported that HPV16 sublineages and single nucleotide polymorphism (SNP), especially the E6 and E7, are associated with the disease status of HPV persistent infection [24–25]. Studies have shown that the increased frequency of HPV 16 E6 gene variations, such as R10G and L83V, is associated to the elevated risk of cervical carcinoma [12, 17, 23, 26–28]. The polymorphisms of HPV L1 gene affect the generation of neutralization antibody of different binding affinities [29].

It was reported that HPV16 is the primary genotype in Henan province, which is located in central China [16]. However, HPV16 sublineages and the nucleotides variations of HPV16 in Henan province are rarely characterized. Principal objective of the present study is to identify the sublineages and analyze the mutations of HPV16 L1, E6 and E7 nucleotides in central China. Investigate the sublineages of HPV16 would assist in the elucidation of the intrinsic geographical relationships. Variations of HPV16 gene may be beneficial for better understanding the functions of HPV genes on oncogenic and the design of HPV vaccine.

Materials And Methods

Sample Collection

From May 2019 to May 2021, cervical swabs from females who attended in the gynecological outpatient in four cities (including Jiyuan, Luohe, Luoyang and Zhengzhou) in Henan province, which are located in central China, were collected. After routine cytology examination, the specimens were subjected to HPV genotyping by flow-through hybridization and gene chip (Hybribio Limited Corporation, Guangdong, China) according to the manufacturer's instruction. The single infected HPV16 positive samples were processed for the variant analysis by sequencing. The study protocol was approved by the institutional ethics committee in the 989 Hospital of Joint Service Support Force of Chinese PLA, Military Training Medical Research Institute of the Whole Army.

HPV genotyping and sequencing

Genomic DNA of the samples were extracted by alkalysis using DNA extraction kit (HybriMax, Hybribio Limited Corporation, Guangdong, China) and then stored at -70°C. The quality and integrity of the extracted DNA were monitored through amplification of the β-globin gene as an internal control. To amplify the full length of the L1, E6 and E7 genes, primers were designed based on published sequences in GenBank (NC 001526). The primers used for the amplification of L1 E6 and E7 genes were shown in Table1. The primers were synthesized by Sangon Biotech, Inc. (Shanghai, China). The PCR reaction volume was 50μl, which included 2μl of template cDNA, 25μl 2×PrimeSTAR Max Premix(Takara Biotechnology Co., LTD, Dalian, China), 2μl of each primer and 17μl of ultrapure water. The PCR program was as follows: initial denaturation step at 94℃ for 10 min; followed by 30 cycles of 95℃ for 30s, 55℃for 30s, 72℃ for 30s, and a final 72℃ extension for 10min. The PCR products were visualized on 1% agarose gels stained with GoldView TM Nucleic Acid Stain. Identified plasmids containing the L1, E6 or E7 genes were used as positive control and the reaction mixture containing no template as negative controls. The targets fragments were then purified using TIANgel Midi Purification Kit (TIANGEN BIOTECH, China) and ligated into p-EASY-Blunt cloning vector (TransGen Biotech, China) according to manufacturer's instruction. The recombinant plasmids were then transformed into Trans1-T1 Phage Resistant Chemically Competent Cells (TransGen Biotech, China) according to manufacturer's protocols. The positive clones containing the recombinant plasmids were sent to Sangon Biotech, Inc. (Shanghai, China) for sequencing.

Molecular characterization and Phylogenetic analysis of HPV16

The variations of the L1, E6 and E7 genes and proteins were compared and numbered with the reference strain HPV16 (GeneBank NC 001526) by DNAStar (Madison, WI, USA). Variants between the studied and reference sequence were noted and the frequency was calculated. Novel variants were identified by the comparisons of the sample sequences with those previously published in GenBank by Blast program.

Phylogenetic tree construction

Phylogenetic analyses of HPV16 L1 and E6-E7 sequences were constructed using the MEGA (version 6.0). A neighbor-joining algorithm was employed and Kimura 2-parameter distance neighbor-joining trees were built with 1000 bootstrapped replicates. The HPV16 reference strains were deposited into NCBI GenBank Database, which represented each lineage [5]. The reference strains include A1 (K02178, KU951179, KX947272 and LC456609), A2 (AF536179), A3 (HQ644236), A4 (AF534061), B1 (AF536180), B2 (HQ644298), C (AF472509), D1 (HQ644257), D2 (AY686579) and D3 (AF402678).

B-cell prediction

B cell epitopes of E6 and E7 sequences were predicted by ABCpred server (https://webs.iiitd.edu.in/raghava/abcpred/ABC_submission.html) at default parameters. The accuracy of ABCpred server that based on recurrent neural network was 65.93%.

Results

Sample characteristics

For the HPV16 L1 and E6-E7 gene, a total of 74 samples (ranging from 16 to 78 years, the median age was 43.7±15.25) were sequencing successfully, including Jiyuan city (9), Luohe city (12), Luoyang city (25) and Zhengzhou city (28). Among of the seventy-four L1 and E6-E7 gene sequences, forty-two have the identical L1 sequences and forty-five have the same E6-E7 sequences, so there were thirty-two and twenty-nine patterns for HPV16 L1 gene (16HNL01-16HNL32) and E6-E7 gene (16HNE01-16HNE22), respectively. The HPV16 L1, E6 and E7 gene sequences were submitted to GenBank and gained the accession number MZ546218-MZ546224, MZ546225-MZ546256 and MZ546257-MZ546285, separately.

L1 gene nucleotide variations and amino acid mutational analysis

Twenty-six novel L1 gene sequences were determined by blast analysis and the 16HNL14 were the highest rate (5/20) (Table 2). Twelve (12/74) HPV16 L1 sequences like 16HNL05 existed and were the most common HPV16 pattern in this study. The thirty-two HPV16 L1 gene shared 99.6%-99.9% identities with the reference sequence (NC 001526) and the variation sites and frequencies of HPV16 L1 gene are shown in Table2. A total of twenty-eight variations (1.8%, 28/1596) were identified in the 1596bp L1 gene. Specially, 18/28 (64.3%) were synonymous mutations and 10/28 (35.7%) were non-synonymous mutations. The nt 874 variation site (A-G) was found in all of 74 samples and brought the amino acid

change from threonine to alanine (T292A). The other high frequent variation sites were A1107C (32/74, 43.2%) and G1500A (50/74, 67.6%) in L1 gene.

E6-E7 gene nucleotide variations and amino acid mutational analysis

Twenty-two novel variations E6-E7 gene sequences were identified by blast comparison. The 16HNE03 (15/74, 20.3%) were the most common HPV16 pattern (Table 3). Compared with the HPV16 reference sequence (NC 001526), the identities of the 29 HPV16 E6-E7 genes were from 99.4–100%. A summary of HPV16 E6-E7 nucleotide sequence variations and frequencies are shown in Table 3. Twenty-five nucleoid changes were found in E6-E7 gene (16 mutations (3.4%, 16/477) in E6 gene, 9 mutations (3.0%, 9/297) in E7 gene), of which 7/25 (28.0%) were synonymous mutations and 18/25 (72.0%) were non-synonymous. The most frequently observed non-synonymous mutations were D32E (29/74, 39.2%) and L90V (5/74, 6.8%) in E6 gene, N29S (27/74, 36.5%) in E7 gene.

Phylogenetic analysis

Phylogenetic analysis based on the full length of L1 genes is shown (Fig. 1). All of the seventy-four L1 sequences, 56.8% (42/74) were A1 sublineage, 37.8% (28/74) were A4 sublineage, 4.1% (3/74) were A3 and 1.4% (1/74) were A2.

Phylogenetic tree of the E6-E7 sequences of HPV16 can be seen (Fig. 2). Among the seventy-four E6-E7 sequences, 37.8% (28/74) were A1 sublineage, 36.5% (27/74) were A4 sublineage, 18.9% (14/74) were A3 and 6.8% (5/74) were A2.

B cell epitopes

The B cell epitopes of E6 and E7 sequences predicted by ABCpred server (only score≥0.80) are summarized in Table4 and Table5. Eight high score epitopes were obtained for E6 sequence and four for E7 sequence. According to the rank, the most potent epitopes were $_{88-103}$YSLYGTTLEQQYNKPL for E6 and $_{29-44}$DGPAGQAEPDRAHYNI for E7. A new epitope, $_{80-95}$ISEYRHYCYSVYGTTL was identified in the E6 variants. Some amino acid mutations have an effect on the score, for example, L90V, R62K and R142Q in E6 and S63F in E7 decreased the score.

Discussion

Cervical cancer is the fourth most common malignancy in females around world [1]. Persistent high-risk HPVs (eg. HPV16, HPV18, et al) infections play an important role in the development of cervical cancer. The prevalence of HPV infection among the females in Henan province had been evaluated and HPV16 is the most predominant genotype [19, 30]. The L1 protein is the major capsid protein and able to induce immune response [29]. Phylogenetic distance and amino variations of the L1 protein influence the immune efficiency of HPV vaccines [10, 31]. The uncontrolled expression of E6 and E7 proteins inactivates the p53 and pRb tumor suppressor proteins and is associated with the HPV persist infection [32]. HPV variants and nucleotide mutations have been suggested to affect the oncogenic potential of HPV persistent infection [23–25, 33]. Sublineages and variations of HPV16 genes in Henan province have not been reported. Thus, HPV16 L1, E6 and E7 sequences were selected to study lineage phylogeny and the genetic polymorphisms.

In the present study, the sublineages of HPV16 among females in Henan province were investigated by sequencing the L1 gene and phylogenic analysis. HPV16 A1-A4 sublineages were all present and the A1 was the most common sublineage in Henan province. The A4 sublineage was the highest in other areas in China, such as Beijing city, Zhejiang and Yunnan province [34–37]. The discrepancies were perhaps due to the HPV geography or ethnicity characteristics [37]. Compared with A1-3 sublineages, it was reported the A4 sublineage are associated with more severity disease status in Chinese females and higher risk of cancer [23, 24, 38, 39].

Compared with the reference amino acid, ten non-synonymous mutations were found in HPV16 L1 protein. The amino mutations N207T (7/74) and T292E (74/74) were also found in other provinces such as Shanghai and Sichuan [37, 40, 41]. Six synonymous mutations in L1 gene, including G1500A (50/74, 67.6%), A1107C (32/74, 43.2%), T987C (12/74, 16.2%), A945G (7/74, 9.5%) and C1266T (6/74, 8.1%) and A237G (3/74, 4.1%) were also reported in another province [41].

Fourteen non-synonymous mutations were found in HPV16 E6 gene and the amino mutations D32E (29/74) and L90V (5/74) were found to have higher mutations rates. It was reported that the D32E mutation was the diagnostic E6 polymorphism for A4 sublineage and L90V was the most common polymorphism within A1–3 sublineages [36]. In our study, the amino mutations D32E (29/74) were found in all of the A4 sublineage and L90V (5/74) were in the A1–3 sublineages. Although the D32E mutation in E6 protein did not induce the degradation of p53 with a higher level compared with the reference HPV16, the gene variation altered the other gene profiles [42, 43]. It was suggested that the D32E amino mutation had a significant correlation with the persistent HPV16 infection in females [37]. Variation of L90V may have discrepant effect on the progression of HPV infection, which depends on the population [44]. The amino mutations L90V (5/74) has no significant association with cervical cancer in Chinese females [36]. Only four non-synonymous mutations were in the E7 gene, which showed the E7 gene was greater stability than E6 gene. The N29S (27/74) was the predominant variant, which was consistent with previous study [41, 45]. The N29S is located in the domain that plays an important role in the binding of E7 protein with the retinoblastoma suppressor protein (pRB) and the complex formation of E7-pRB that is associated to cervical carcinoma [46]. Furthermore, it suggested that the amino 29 was involved in the host immune recognition and lesion progression [47].

In summary, the sublineage of HPV16 prevalent in the central China was first reported to be A1. The HPV16 sublineage in Henan province would be beneficial for charactering the HPV gene polymorphism in China. Phylogeny trees based on the HPV16 L1 and E6-E7 sequences were constructed. Variations of the HPV16 L1, E6 and E7 were identified and some novel variants were found. However, the relationships between the variations and their effects on the protein function should be further investigated. The HPV16 L1 sequences may assistant on the improvement of HPV vaccines. Associations between the HPV16 E6-E7 sequences and the regression of HPV-related lesion in cervical cancer need to be studied further.

Declarations

Acknowledgments

The authors are grateful to the staff at the Department of Gynecological Outpatient Clinic, the 989th hospital of the People's Liberation Army (PLA), Luoyang, Henan, China; Department of Clinical Laboratory, the Central Hospital of Luohe City, Luohe, Henan, China; Department of Clinical Laboratory, Zhengzhou Central Hospital, Zhengzhou, Henan, China; Department of Clinical Laboratory, Jiyuan Shi People's Hospital, Jiyuan, Henan, China, for the sample collection.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.CA Cancer J Clin68:394–424
2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.Int J Cancer136:E359-386
3. Lu Y, Li P, Luo G, Liu D, Zou H (2020) Cancer attributable to human papillomavirus infection in China: Burden and trends.Cancer126:3719–3732
4. Bzhalava D, Eklund C, Dillner J (2015) International standardization and classification of human papillomavirus types.Virology476:341–344
5. Burk RD, Harari A, Chen Z (2013) Human papillomavirus genome variants.Virology445:232–243
6. Scheffner M, Romanczuk H, Munger K, Huibregtse JM, Mietz JA, Howley PM (1994) Functions of human papillomavirus proteins.Curr Top Microbiol Immunol186:83–99
7. Seedorf K, Krammer G, Durst M, Suhai S, Rowekamp WG (1985) Human papillomavirus type 16 DNA sequence.Virology145:181–185
8. Bishop B, Dasgupta J, Klein M, Garcea RL, Christensen ND, Zhao R, Chen XS (2007) Crystal structures of four types of human papillomavirus L1 capsid proteins: understanding the specificity of neutralizing monoclonal antibodies.J Biol Chem282:31803–31811
9. Mo X, Gai Tobe R, Wang L, Liu X, Wu B, Luo H, Nagata C, Mori R, Nakayama T (2017) Cost-effectiveness analysis of different types of human papillomavirus vaccination combined with a cervical cancer screening program in mainland China.BMC Infect Dis17:502
10. Bogaards JA, van der Weele P, Woestenberg PJ, van Benthem BHB, King AJ (2019) Bivalent Human Papillomavirus (HPV) vaccine effectiveness correlates with phylogenetic distance from HPV vaccine types 16 and 18.J Infect Dis220:1141–1146
11. Yao Y, Huang W, Yang X, Sun W, Liu X, Cun W, Ma Y (2013) HPV-16 E6 and E7 protein T cell epitopes prediction analysis based on distributions of HLA-A loci across populations: an in silico approach.Vaccine31:2289–2294
12. Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, Snijders PJ, Meijer CJ, International Agency for Research on Cancer Multicenter Cervical Cancer Study G (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer.N Engl J Med348:518–527
13. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, Tous S, Felix A, Bravo LE, Shin HR, Vallejos CS, de Ruiz PA, Lima MA, Guimera N, Clavero O, Alejo M, Llombart-Bosch A, Cheng-Yang C, Tatti SA, Kasamatsu E, Iljazovic E, Odida M, Prado R, Seoud M, Grce M, Usubutun A, Jain A, Suarez GA, Lombardi LE, Banjo A, Menendez C, Domingo EJ, Velasco J, Nessa A, Chichareon SC, Qiao YL, Lerma E, Garland SM, Sasagawa T, Ferrera A, Hammouda D, Mariani L, Pelayo A, Steiner I, Oliva E, Meijer CJ, Al-Jassar WF, Cruz E, Wright TC, Puras A, Llave CL, Tzardi M, Agorastos T, Garcia-Barriola V, Clavel C, Ordi J, Andujar M, Castellsague X, Sanchez GI, Nowakowski AM, Bornstein J, Munoz N, Bosch FX, Retrospective International S, Group HPVTTS (2010) Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study.Lancet Oncol11:1048–1056
14. Yan X, Huang Y, Zhang M, Hu X, Li K, Jing M (2020) Prevalence of human papillomavirus infection and type distribution among Uyghur females in Xinjiang, northwest China.Oncol Lett20:25
15. W ZGDZXY, Z QFXBL H, S Y (2020) Human papillomavirus infection rates before and after the introduction of prophylactic vaccines in Kunming, Yunnan, China.Indian J Med Microbiol38:66–71
16. Liu J, Ma S, Qin C, Zheng S, Chen Z, Huang Y, Xiong J, Huo Y (2020) Prevalence and genotype distribution of human papillomavirus in Zhengzhou, China, in 2016.Arch Virol165:731–736
17. Chen W, Zhang X, Molijn A, Jenkins D, Shi JF, Quint W, Schmidt JE, Wang P, Liu YL, Li LK, Shi H, Liu JH, Xie X, Niyazi M, Yang P, Wei LH, Li LY, Li J, Liu JF, Zhou Q, Hong Y, Li L, Li Q, Zhou HL, Bian ML, Chen J, Qiao YL, Smith JS (2009) Human papillomavirus type-distribution in cervical

cancer in China: the importance of HPV 16 and 18.Cancer Causes Control20:1705–1713

18. Ge Y, Zhong S, Ren M, Ge Y, Mao Y, Cao P (2019) Prevalence of human papillomavirus infection of 65,613 women in East China.BMC Public Health19:178
19. Wang XC, Sun LQ, Ma L, Li HX, Wang XL, Wang X, Yun T, Meng NL, Lv DL (2014) Prevalence and genotype distribution of human papillomavirus among women from Henan, China.Asian Pac J Cancer Prev15:7333–7336
20. Yamada T, Manos MM, Peto J, Greer CE, Munoz N, Bosch FX, Wheeler CM (1997) Human papillomavirus type 16 sequence variation in cervical cancers: a worldwide perspective.J Virol71:2463–2472
21. Fang L, Lin X, Yang Y, Song Z, Ding X, Tan L, Gao P (2020) Genetic variability, phylogeny and functional implication of the long control region in human papillomavirus type 16, 18 and 58 in Chengdu, China.Virol J17:106
22. Zhe X, Xin H, Pan Z, Jin F, Zheng W, Li H, Li D, Cao D, Li Y, Zhang C, Fu S, Shao R, Pan Z (2019) Genetic variations in E6, E7 and the long control region of human papillomavirus type 16 among patients with cervical lesions in Xinjiang, China.Cancer Cell Int19:65
23. Sun Z, Lu Z, Liu J, Wang G, Zhou W, Yang L, Liu C, Wang B, Ruan Q (2013) Genetic variations of E6 and long control region of human papillomavirus type 16 from patients with cervical lesion in Liaoning, China.BMC Cancer13:459
24. Mirabello L, Yeager M, Cullen M, Boland JF, Chen Z, Wentzensen N, Zhang X, Yu K, Yang Q, Mitchell J, Roberson D, Bass S, Xiao Y, Burdett L, Raine-Bennett T, Lorey T, Castle PE, Burk RD, Schiffman M (2016) HPV16 Sublineage Associations With Histology-Specific Cancer Risk Using HPV Whole-Genome Sequences in 3200 Women.J Natl Cancer Inst108
25. van der Weele P, Meijer C, King AJ (2017) Whole-Genome Sequencing and Variant Analysis of Human Papillomavirus 16 Infections.J Virol91
26. Yu LL, Kang LN, Zhao FH, Lei XQ, Qin Y, Wu ZN, Wang H, Chen W, Qiao YL (2016) Elevated Expression of Human Papillomavirus-16/18 E6 Oncoprotein Associates with Persistence of Viral Infection: A 3-Year Prospective Study in China.Cancer Epidemiol Biomarkers Prev25:1167–1174
27. Zehbe I, Voglino G, Delius H, Wilander E, Tommasino M (1998) Risk of cervical cancer and geographical variations of human papillomavirus 16 E6 polymorphisms.Lancet352:1441–1442
28. Chan PK, Lam CW, Cheung TH, Li WW, Lo KW, Chan MY, Cheung JL, Xu LY, Cheng AF (2002) Human papillomavirus type 16 intratypic variant infection and risk for cervical neoplasia in southern China.J Infect Dis186:696–700
29. Varsani A, Williamson AL, Jaffer MA, Rybicki EP (2006) A deletion and point mutation study of the human papillomavirus type 16 major capsid gene.Virus Res122:154–163
30. Chen G, Zheng P, Gao L, Zhao J, Wang Y, Qin W (2020) Prevalence and genotype distribution of human papillomavirus in women with cervical cancer or cervical intraepithelial neoplasia in Henan province, central China.J Med Virol
31. Godi A, Boampong D, Elegunde B, Panwar K, Fleury M, Li S, Zhao Q, Xia N, Christensen ND, Beddows S (2020) Comprehensive Assessment of the Antigenic Impact of Human Papillomavirus Lineage Variation on Recognition by Neutralizing Monoclonal Antibodies Raised against Lineage A Major Capsid Proteins of Vaccine-Related Genotypes. J Virol 94
32. Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, Stanley MA (2012) The biology and life-cycle of human papillomaviruses.Vaccine30 Suppl 5:F55-70
33. Xi LF, Schiffman M, Koutsky LA, Hughes JP, Winer RL, Mao C, Hulbert A, Lee SK, Shen Z, Kiviat NB (2014) Lineages of oncogenic human papillomavirus types other than type 16 and 18 and risk for cervical intraepithelial neoplasia.J Natl Cancer Inst106
34. Zhou Z, Yang H, Yang L, Yao Y, Dai S, Shi L, Li C, Yang L, Yan Z, Yao Y (2019) Human papillomavirus type 16 E6 and E7 gene variations associated with cervical cancer in a Han Chinese population.Infect Genet Evol73:13–20
35. Ding T, Wang X, Ye F, Cheng X, Lu W, Xie X (2010) Distribution of human papillomavirus 16 E6/E7 variants in cervical cancer and intraepithelial neoplasia in Chinese women.Int J Gynecol Cancer20:1391–1398
36. Hang D, Yin Y, Han J, Jiang J, Ma H, Xie S, Feng X, Zhang K, Hu Z, Shen H, Clifford GM, Dai M, Li N (2016) Analysis of human papillomavirus 16 variants and risk for cervical cancer in Chinese population.Virology488:156–161
37. Yang B, Zhang L, Zhang A, Zhou A, Yuan J, Wang Y, Sun L, Cao H, Zheng W (2019) Variant sublineages of human papillomavirus type 16 predispose women to persistent infection characterized by a sequence analysis of the E6, L1, and LCR regions.Int J Clin Exp Pathol12:337–343
38. Zhang J, Wang T, Han M, Yang ZH, Liu LX, Chen Y, Zhang L, Hu HZ, Xi MR (2010) Variation of human papillomavirus 16 in cervical and lung cancers in Sichuan, China.Acta Virol54:247–253
39. Freitas LB, Chen Z, Muqui EF, Boldrini NA, Miranda AE, Spano LC, Burk RD (2014) Human papillomavirus 16 non-European variants are preferentially associated with high-grade cervical lesions.PLoS One9:e100746
40. Yue Y, Yang H, Wu K, Yang L, Chen J, Huang X, Pan Y, Ruan Y, Zhao Y, Shi X, Sun Q, Li Q (2013) Genetic variability in L1 and L2 genes of HPV-16 and HPV-58 in Southwest China.PLoS One8:e55204
41. Cao M, Chenzhang Y, Ding X, Zhang Y, Jing Y, Chen Z (2016) Genetic variability and lineage phylogeny of human papillomavirus type-16 and -53 based on the E6, E7, and L1 genes in Southwest China.Gene592:49–59

42. Yi JW, Jang M, Kim SJ, Kim SS, Rhee JE (2013) Degradation of p53 by natural variants of the E6 protein of human papillomavirus type 16.Oncol Rep29:1617–1622
43. Jang M, Rhee JE, Jang DH, Kim SS (2011) Gene expression profiles are altered in human papillomavirus-16 E6 D25E-expressing cell lines.Virol J8:453
44. Cornet I, Gheit T, Iannacone MR, Vignat J, Sylla BS, Del Mistro A, Franceschi S, Tommasino M, Clifford GM (2013) HPV16 genetic variation and the development of cervical cancer worldwide.Br J Cancer108:240–244
45. Shang Q, Wang Y, Fang Y, Wei L, Chen S, Sun Y, Li B, Zhang F, Gu H (2011) Human papillomavirus type 16 variant analysis of E6, E7, and L1 [corrected] genes and long control region in [corrected] cervical carcinomas in patients in northeast China.J Clin Microbiol49:2656–2663
46. Jones RE, Wegrzyn RJ, Patrick DR, Balishin NL, Vuocolo GA, Riemen MW, Defeo-Jones D, Garsky VM, Heimbrook DC, Oliff A (1990) Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein.J Biol Chem265:12782–12785
47. Stephen AL, Thompson CH, Tattersall MH, Cossart YE, Rose BR (2000) Analysis of mutations in the URR and E6/E7 oncogenes of HPV 16 cervical cancer isolates from central China.Int J Cancer86:695–701

Tables

Table 1. Primers used for the amplification of HPV16 L1 E6 and E7 genes.

Primer name	Start codon	Sequence 5-3	Amplicon size(bp)
HPV16 L1 1F	4636	GACCAAGCTCCTTCATTAATTCCT	1051
HPV16 L1 1R	5686	GGCATCAGAGGTAACCATAGAAC	
HPV16 L1 2F	5393	GCTATGGACTTTACTACATTACAGGC	971
HPV16 L1 2R	6363	TTTACAAGCACATACAAGCACATA	
HPV16 E6 F	7119	TTATGCACCAAAAGAGAACTGCA	502
HPV16 E7 F	7595	GCTGTAATCATGCATGGAGATACACCT	316
HPV16 E7 R	4	GCAGGATCAGCCATGGTAGATTAT	

Table 2. Nucleotide sequence mutations of HPV16 L1 genes

nt	Domain: HPV16 L1 sequence																													n	Sublineage
			2	2	2	3	3	4	6	7	8	8	9	9	9	1	1	1	1	1	1	1	1	1	1	1	1	1		
		4	2	3	5	0	3	4	2	9	3	7	3	4	8	0	0	1	2	2	2	2	2	3	3	3	4	5		
	8	9	8	7	1	6	4	1	0	8	2	4	0	5	7	3	8	7	7	3	6	8	4	9	9	2	3	0		
NC 001526	T	G	T	A	A	T	C	T	A	A	A	A	A	A	T	A	A	A	A	A	C	G	C	T	A	A	T	G		Protype
16HNL01												G				G												A	1	A4
16HNL02												G						C			T								1	A1
16HNL03												G		G														A	6	A4
16HNL04			C		G							G			C			C										A	2	A1
16HNL05												G																A	12	A4
16HNL06								G				G									T						G	A	1	A3
16HNL07			C									G			C			C										A	2	A1
16HNL08				G								G																	2	A1
16HNL09												G																	6	A1
16HNL10												G									T					G		A	1	A3
16HNL11												G						T			T								1	A1
16HNL12												G	G															A	1	A4
16HNL13												G		G				C											1	A1
16HNL14												G			C			C										A	5	A1
16HNL15												G						C											7	A1
16HNL16				G								G			C			C										A	1	A1
16HNL17											C	G																	1	A1
16HNL18	C											G			C			C										A	1	A1
16HNL19						G			C			G			C			C										A	1	A1
16HNL20												G									T							A	1	A3

nt	Domain: HPV16 L1 sequence																												n	Sublineage
			2	2	2	3	3	4	6	7	8	8	9	9	9	10	10	11	12	12	12	12	12	13	13	13	14	15		
		4	2	3	5	0	3	4	2	9	3	7	3	4	8	2	9	0	1	3	6	7	9	5	8	9	1	0		
	8	9	8	7	1	6	4	1	0	8	2	4	0	5	7	3	8	7	7	3	6	8	4	9	9	2	3	0		
NC_00152(	T	G	T	A	A	T	C	T	A	A	A	A	A	A	T	A	A	A	A	A	C	G	C	T	A	A	T	G		Protype
16HNL21									C			G									T							A	1	A4
16HNL22												G					T	C										A	2	A1
16HNL23												G					T	C				A						A	1	A1
16HNL24										C		G																A	1	A4
16HNL25									C			G																A	5	A4
16HNL26												G						C										A	3	A1
16HNL27							T					G																A	1	A4
16HNL28		A										G					T	C							C			A	1	A1
16HNL29												G						C					T						2	A1
16HNL30												G												G					1	A1
16HNL31												G						C		T									1	A1
16HNL32												G						C	C										1	A2
reference a	V	D	F	K	N	H	P	P	N	E	R	T	A	S	D	R	T	S	K	A	S	E	L	A	K	E	T	L		
						11	11	11	12	12	12	12	13	13	13	13	13	13	14	14	14	14	14	14	14	14	14	15		
aa position		17	76	79	84	02	12	47	70	66	78	92	10	15	92	14	66	69	60	11	22	62	23	53	63	74	10			
	3	7	6	9	4	2	2	7	7	6	8	2	0	5	9	1	6	9	6	1	2	6	2	3	3	4	1	0		
aa mutations	A	N			S	Q	S		T	D		A							T						N					
secondary structure	Alpha helix					Sheet					Alpha helix							Sheet		Alpha helix		Alpha helix		Alpha helix						

Note: numbers of the variation in the domain of the L1 genes were statistics in bracket. Nucleotide sequence data reported are available in the GenBank databases under the accession numbers MZ546218-MZ546224.

Table 3. Nucleotide sequence mutations of HPV16 E6-E7 genes

nt	Domain: HPV16 E6 sequence																Domain: HPV16 E7 sequence								n	Sublineage
	1	1	4	8	9	9	10	10	10	10	13	18	12	16	13	12	14	14	16	14	16	19	14	14		
	2	3	9	6	4	6	0	1	2	6	7	5	8	8	4	5	6	7	6	9	0	0	9	3	6	
NC001526	G	A	A	C	G	T	A	T	A	G	G	G	T	T	A	G	A	A	G	C	T	C	A	T	T	Protype
16HNE01						G												G						C	C	7	A4
16HNE02																			A							7	A3
16HNE03						G												G							C	#	A4
16HNE04												A									C					2	A1
16HNE05															T				A							1	A3
16HNE06																					C					6	A1
16HNE07						G							G					G						C	C	1	A4
16HNE08								C														T				1	A1
16HNE09	A													G			C									1	A2
16HNE10						G												G		T				C	C	1	A4
16HNE11	A															A			A	T						2	A3
16HNE12						G												G		T					C	2	A4
16HNE13																						T				2	A1
16HNE14														G						T						1	A2
16HNE15																			A	T						2	A3
16HNE16																				T	C					1	A1
16HNE17	A					A													A							1	A3
16HNE18	A	G				A													A							1	A3
16HNE19					A																C					1	A1
16HNE20				G							A			G			C									1	A2

nt	Domain: HPV16 E6 sequence																Domain: HPV16 E7 sequence									n	Sublineage
	1	1	4	8	9	9	1	1	1	1	1	1	2	2	3	4	6	6	6	7	7	7	8	8	8		
	0	0	0	0	3	8	2	6	3	2	4	4	6	4	6	9	1	4	4								
	2	3	9	6	4	6	0	1	2	6	7	5	8	8	4	5	6	7	6	9	0	0	9	3	6		
NC001526	G	A	A	C	G	T	A	T	A	G	G	G	T	T	A	G	A	A	G	C	T	C	A	T	T		Protype
16HNE21					T																C					1	A1
16HNE22										C																2	A1
16HNE23																										9	A1
16HNE24	A						C							G												1	A2
16HNE25	A																									1	A1
16HNE26				G										G			C									1	A2
16HNE27																							G			1	A1
16HNE28									G																	1	A1
16HNE29			C			G												G							C	1	A4
reference aa	K	R	R	T	D	D	I	I	I	E	R	R	F	L	N	R	N	N	E	S	L	R	T	C	S		
															1	1											
			1	2	3	3	3	3	3	3	4	6	7	9	1	4	2	2	3	6	6	7	8	9	9		
	4	5	7	9	2	2	4	4	4	6	6	2	6	0	2	2	9	9	5	3	7	7	6	4	5		
aa mutations		G		S	N/Y	E	L	T	M	Q	Q	K	L	V	Y	Q	H	S		F		C					
secondary structure					Alpha helix	Alpha helix	Sheet	Sheet	Sheet	Sheet	Alpha helix	Sheet	Alpha helix	Sheet			Alpha helix	Alpha helix		Sheet	Sheet	Alpha helix					
Frenquency	7	1	1	2	2	29	1	1	1	2	1	2	1	5	1	2	3	27	14	9	11	3	1	9	27	#	

Note: numbers of the variation in the domain of the E6-E7 genes were statistics in bracket. Nucleotide sequence data reported are available in the GenBank databases under the accession numbers MZ546225-MZ546256 (E6) and MZ546257-MZ546285 (E7).

Table 4
Predicted B-cell epitopes of the HPV16 E6

Reference sequence				E6 Variants sequence				
Rank	Sequence	Start position	Score	Rank	Sequence	Start position	Score	No. sequence
1	YSLYGTTLEQQYNKPL	88	0.87	1	YSLYGTTLEQQYNKPL	88	0.87	16HNE01-08/10-13/15-19/21-23/25/27-29
				1	YSVYGTTLEQQYNKPL	88	0.86	16HNE09/14/20/24/26
2	RDLCIVYRDGNPYAVC	55	0.86	1	RDLCIVYRDGNPYAVC	55	0.86	16HNE01-03/05-29
				3	DLCIVYKDGNPYAVCD	56	0.85	HNE04
2	FHNIRGRWTGRCMSCC	132	0.86	1	FHNIRGRWTGRCMSCC	132	0.86	16HNE01-10/12-29
				7	FHNIRGRWTGQCMSCC	132	0.81	HNE11
3	TAMFQDPQERPRKLPQ	6	0.85	2	TAMFQDPQERPRKLPQ	6	0.85	16HNE01-29
4	YRDGNPYAVCDKCLKF	61	0.84	4	YRDGNPYAVCDKCLKF	61	0.84	HNE01-03/05-29
5	LKFYSKISEYRHYCYS	74	0.83	5	LKFYSKISEYRHYCYS	74	0.83	HNE01-06/08-29
				2	LKLYSKISEYRHYCYS	74	0.86	16HNE07
				5	ISEYRHYCYSVYGTTL	80	0.82	HNE09/14/20/24/26
5	RWTGRCMSCCRSSRTR	138	0.83	5	RWTGRCMSCCRSSRTR	138	0.83	HNE01-29
6	YAVCDKCLKFYSKISE	67	0.8	6	YAVCDKCLKFYSKISE	67	0.82	HNE01-06/08-29
Note: Sequences with amino acids change were highlight in red.								

Table 5
Predicted B-cell epitopes of the HPV16 E7

Reference sequence				E7 Variants sequence				
Rank	Sequence	Start position	Score	Rank	Sequence	Start position	Score	No. sequence
1	DGPAGQAEPDRAHYNI	39	0.85	1	DGPAGQAEPDRAHYNI	39	0.85	16HNE01-29
2	DRAHYNIVTFCCKCDS	48	0.82	2	DRAHYNIVTFCCKCDS	48	0.82	16HNE01-09/13/17-29
				4	DRAHYNIVTFCCKCDF	48	0.8	HNE10-12/14-16
2	HGDTPTLHEYMLDLQP	2	0.82	2	HGDTPTLHEYMLDLQP	2	0.82	16HNE01-29
3	EQLNDSSEEEDEIDGP	26	0.81	3	EQLNDSSEEEDEIDGP	26	0.81	HNE02/04-06/08/11/13-19/21-25/27-28
				2	EQLSDSSEEEDEIDGP	26	0.83	16HNE01/03/07/10/12/29
				2	HDSSEEEDEIDGPAGQ	29	0.82	16HNE09/20/26
Note: Sequences with amino acids change were highlight in red.								

Figures

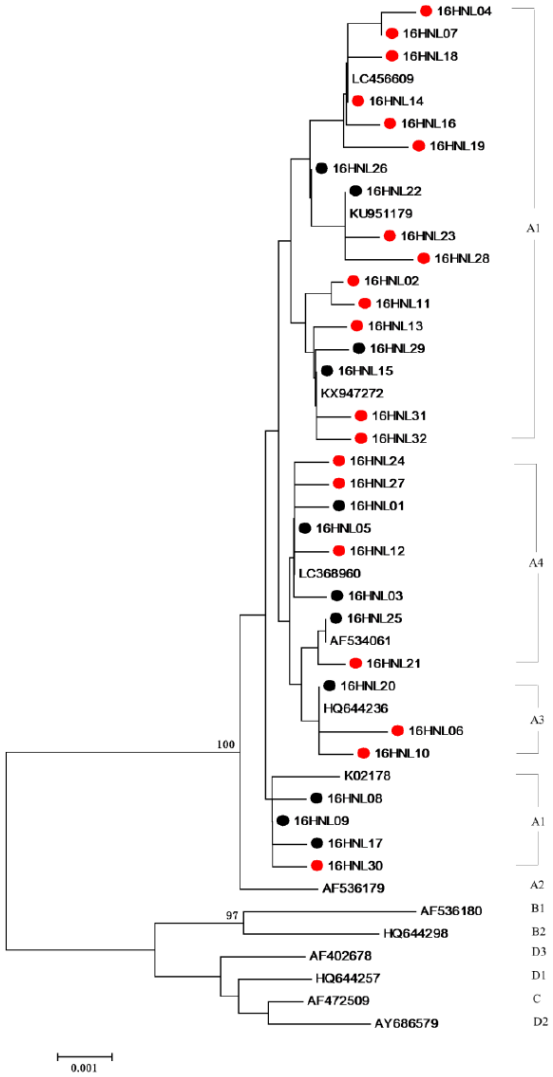

Figure 1

Neighbor joining phylogenetic tree generated using nucleotide sequences of the HPV16 L1 gene. Legend: Study sequence are labeled in dots and novel sequences are highlighted in red dots, others without dots are reference strain, including: A1 (K02178, KU951179, KX947272 and LC456609), A2 (AF536179), A3 (HQ644236), A4 (AF534061 and LC368960), B1 (AF536180), B2 (HQ644298), C (AF472509), D1 (HQ644257), D2 (AY686579), D3 (AF402678). Phylogenetic trees were constructed by the Neighbor-Joining method and the Kimura 2-parameter model by MEGA 6.0 package. Only bootstrap values above 70% are displayed in the branches.

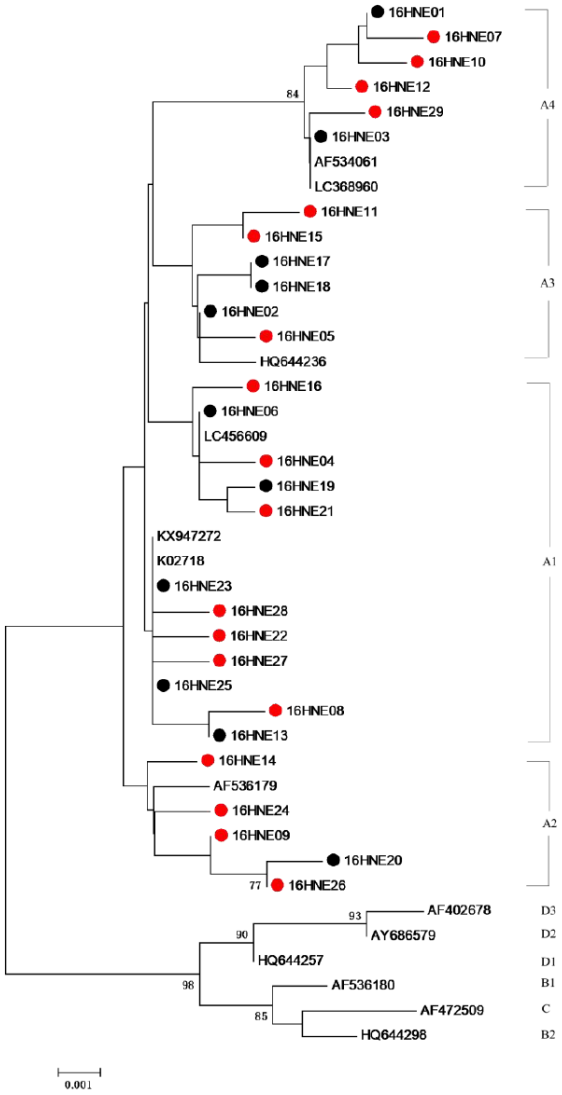

Figure 2

Neighbor joining phylogenetic tree generated using nucleotide sequences of the HPV16 E6-E7 gene. Legend: Study sequence are labeled in dots and novel sequences are highlighted in red dots, others without dots are reference strain, including: A1 (K02178, KU951179, KX947272 and LC456609), A2 (AF536179), A3 (HQ644236), A4 (AF534061 and LC368960), B1 (AF536180), B2 (HQ644298), C (AF472509), D1 (HQ644257), D2 (AY686579), D3 (AF402678). Phylogenetic trees were constructed by the Neighbor-Joining method and the Kimura 2-parameter model by MEGA 6.0 package. Only bootstrap values above 70% are displayed in the branches.