GENETIC ANALYSIS OF HUMAN PAPILLOMAVIRUS IN A COHORT OF WOMEN IN ROUTINE CARE IN NORTHERN SOUTH AFRICA

BY

RIKHOTSO RIXONGILE RHENNY

A dissertation submitted in fulfilment of the requirements for the award of the degree of Masters in Microbiology

To

The Department of Microbiology, School of Mathematical and Natural Sciences, University of Venda

South Africa

Supervisor: Prof Bessong Pascal Obong

(University of Venda)

Co-Supervisor: Dr Lufuno Grace Mavhandu-Ramarumo

(University of Venda)

2019

DECLARATION

I, Rikhotso Rixongile Rhenny, hereby declare that this dissertation for the award of Master of Science degree in Microbiology at the University of Venda is my own work. It has not been submitted before for any degree examination at this or any other University. It is my own work at execution and all the reference materials contained are therein have been duly acknowledged.

Signature……..…… Date…….……

ACKNOWLEDGEMENTS

To God be the glory, I would never have made it this far if it wasn't of your unmerited favour. I am so much thankful for granting me the gift of life, good health and strength to do this work.

I would like to acknowledge the Department of Microbiology, without your permission to enrol for my post graduate studies this dissertation wouldn't have been possible

To my supervisor Prof Pascal Obong Bessong, throughout this journey you have not been anything less but a great teacher, guider, mentor, leader, and an advisor. Your words of encouragement, valuable suggestions and comments when I feel like I cannot do this work made this work to come into completion today. Thank you so much Prof for not giving up on me, I will forever run with the knowledge, skills and the words of wisdom you graced me with. You always say to me I quote "You need to deliver on what is expected of you, self-pity will not help you".

I extend my gratitude to my Co-Supervisor Dr Lufuno Grace Mavhandu-Ramarumo for all the contributions you made towards my work, I am very much grateful.

I would like to extend my sincere gratitude to my brother and a friend Rodney Rikhotso; you are that one person who believes in my dreams. I must say your support and prayers made a huge difference in my MSc work.

To the Rikhotso family at large, a warm thank you for the emotional support you gave me. All the late phone calls when I feel like throwing up the towel played a huge role.

To the entire team of HIV/AIDS and Global Health Programme, thank you for all the scientific inputs you contributed towards my work.

I LOVE YOU ALL…….

DEDICATION

I would like to dedicate this dissertation to my late parents Mr Gezani Robert and Mrs Latani Veronica Rikhotso, gone yet not forgotten. May your beautiful souls continue to rest in perfect peace. I love you.

ABSTRACT

BACKGROUND: Human papillomavirus (HPV) is a common sexually transmitted virus known to be a causative agent of cervical cancer (CC), one of the most frequent cancers in women worldwide. HPV is a double stranded DNA virus of approximately 7,900 bp; belonging to *Papillomaviridae* family. To date, about 202 low risk (LR) and high risk (HR) HPV genotypes have been identified. However, available vaccines against HPV infection are designed based on the most common known genotypes. Therefore, it is critical to understand the scope and diversity of HPV genotypes in all geographical locations which can help to inform the design and development of future vaccines.

OBJECTIVE: The objective of this study was to describe the burden and diversity of HPV genotypes in a cohort of women in routine care in northern South Africa.

METHODS: Eighty seven women consented to participate in the study and each provided a specimen for analysis. With the help of qualified health care practitioners, Aptima Cervical Specimen Collection and Transport Kit (Hologic, San Diego, CA) was used to collect cervical specimens from each study participant following the manufacturer's procedure. Total DNA was purified from the cervical pellet using QIAamp DNA mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The purified DNA was then subjected to a single round conventional PCR in a reaction volume of 100 µl to amplify HPV L1 gene comprising of approximately 450 bp. A portion of each PCR amplicon from each participant was denatured, hybridized and genotyped using the Linear Array HPV genotyping Test Kit (Roche Molecular Systems, Inc. Branchburg, NJ USA). The kit is designed to detect 37 HPV genotypes (genotypes 6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73, 81, 82, 83, 84, IS39 and CP6108). To detect the HPV genotypes, the Linear Array (LA) reference guide was used for results interpretation following the manufacturer's instructions. The other portion of each of the amplicons was subjected to next generation sequencing (NGS) using the Illumina MiniSeq platform. Using the Nextera XT DNA Library preparation kit, an initial input of 1ng genomic DNA was tagmented, cleaned up, normalized and pooled. The pooled library was then denatured with 0.1

N NaOH and diluted into a final volume of 500 µl at 1.8 pM then sequenced using the Local Run Manager option following the manufacturer's instructions. The generated sequence data was downloaded into fastaQ format and analysed using Genious 11.0.5 software.

RESULTS: Of the 87 participants, the overall proportion of women harbouring HPV DNA by linear array (LA) PCR was 23% (n=20). Of the 20, 16 (80%) were living with HIV. However, this difference was not significant (p=0.077). Genotyping data generated by Roche LA method was successful for all the 20 positive amplicons. In this study, 27 (73%) of the 37 HPV genotypes incorporated in the Roche Linear Array method were detected. The detected genotypes include: types 84, 83, 81, 73, 72, 71, 70, 69, 68, 66, 62, 61, 59, 54, 53, 52, 51, 45, 42, 39, 35, 26, 18, 16, 6, IS39 and CP6108. Most women (15/20;75%) harboured multiple infections compared to single infection. In terms of genotypes distribution, the most frequent genotypes detected LR HPV types in increasing order of frequency included HPV type 61 and 83 (12%), 62 (36%) and 81 (43%). On the other hand, HPV type 66, 53, 52, 51, 18 and 16 were the most common genotypes detected HR HPV types.

In contrast, although genotyping data was successfully generated from 15 of 20 women (75%), NGS technology was seen to be more sensitive compared to Roche LA method. Nearly all the detected genotypes identified by the commercial kit were detected by NGS. In addition, NGS detected 10 namely: HPV types 11, 31, 33, 40, 55, 56, 58, 64, 67, and 82 that were not detected by the LA yet incorporated in the kit. Moreover, it was observed that NGS identified additional 6 HPV types including HPV types 2, 27, 30, 35, 85 and 102 not incorporated in the Roche LA kit. A similar distribution of HPV multiple infections was observed in the study population, however, high frequency of 93% (14 of 15) was detected by NGS. The proportion of women harbouring one or more of the 22 LR HPV types was 100% (n=15).The most frequent LR genotypes in increasing order of frequency was HPV type 62 and 70 (27%), 6 (40%) and 11 (47%). HPV types 40, 42, 54, 72, 64, and 81 were the least detected genotypes with n=1 (7%) each. Furthermore, the common combination observed among the participants was type 6 and 11. In contrast, the most frequent detected genotypes in the study population by NGS under the HR HPV types in increasing order of frequency include type 35 (21%), 39, 56 and 82 (29%), 68 (36%) and 51 (50%). In addition, HPV types 26, 31, 45, 53, 56, 58 and 66 were the least

detected genotypes n=1 (7%) in the study population. HPV 39 and 68 were observed as the common combination detected under HR HPV types. Following genotyping by LA and NGS, the demographic and clinical data of all the 20 positive subjects by PCR were subjected to statistical analysis to determine the association between HPV positive DNA status and associated risk factors. Smoking status (p=0.000), age at first sexual intercourse (p=0.011), vaccination status (p=0.000), gender of sexual partner (p=0.000), highest level of education (p=0.004), marital status (p=0.008) and number of sexual partners (p=0.000) were found to be having a positive statistical association.

CONCLUSION: Amplification of targeted HPV DNA from cervical specimens demonstrated the presence of HPV infection in the study cohort, with a proportion of 23%. The findings illustrate that there is a diversity of HPV genotypes prevalent in the study population as shown by Roche LA and NGS methods. However, the NGS method was observed to be more sensitive than Roche LA in detecting HPV genotypes. Furthermore, NGS identified 6 additional HPV types not incorporated in the Roche LA. Thus, there are genotypes that may be present in the study population that the Roche commercial kit may fail to detect. Therefore, is it imperative to use both genotyping methods to confirm HPV genotypes.

Key words: Human papillomavirus; Cervical cancer; Genotypes; Linear Array Genotyping; Next generation sequencing; South Africa

TABLE OF CONTENTS

CHAPTER TWO: MATERIALS AND METHODS

CHAPTER THREE: RESULTS

CHAPTER FOUR: DISCUSSION

LIST OF ABBREVIATIONS

AA	Asia-American
AE	Elution buffer
Af1	African-1
Af2	African-2
AS	Asia
AIDS	Acquired immunodeficiency syndrome
A_E	Early Polyadenylation
A_L	Late Polyadenylation
ARV	Anti-retroviral therapy
ASCUS	Atypical squamous cells of undetermined significance
bp	Base pairs
CC	Cervical cancer
CIN	Cervical intraepithelial neoplasia
DF	Donald Fraser
DNA	Deoxyribonucleic acid
ds	Double stranded
E1-E7	Early regions
E	European
et al	And others
EV	Epidermodysplasia verruciformis

FDA	Food and drug administration
HIV	Human immunodeficiency virus
HPV	Human papillomavirus
HR	High-risk Human papillomavirus
ISH	In situ hybridization
L1	Late region (major protein)
LA	Linear Array
L2	Late region (minor protein)
LCR	Long control region
LR	Low-risk
ml	Milliliter
Min	Minute(s)
mRNA	Messenger Ribonucleic acid
n	Number of
N	Normality
NaOH	Sodium hydroxide
ng	Nano gram
nM	Nano molar
ng/ml	Nano gram per milligram
NGS	Next generation sequencing
nm	Nanometer
ORF(s)	Open reading frame(s)
P	p-value
PBS	Phosphate buffer saline

PCR	Polymerase chain reaction
pH	Potential of hydroden
pM	Pico molar
R	Rethabile Community Health Center
RBS	Re-suspension buffer
RPM	Revolution Per Minute
S	Seshego Community Health Center
Sec	Second(s)
SPSS	Software package used for statistical analysis
TH	Thohoyandou Community Health Center
VLPs	Virus like particles
WHO	World health organization

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

Several studies conducted worldwide on human papillomavirus had highlighted a strong relationship between human immunodeficiency virus (HIV), human papilloma virus (HPV) and cervical cancer (CC). Cervical cancer is one of the HIV/AIDS associated malignancy known to be caused by high risk persistent HPV types. A higher burden of CC has been observed in the HIV-infected population compared with the general population (Sheils et al., 2011). However, the introduction of antiretroviral therapy (ART) has prolonged the life span of individuals living with HIV/AIDS (Samji et al., 2013; Wada et al., 2013), leading to an increasing number of people aging with HIV infection (Justice, 2010). Thus, allowing HIV-infected patients to experience life expectancy approaching that of persons without HIV (van Sighem et al., 2010; Nakagawa et al., 2012). An increase in individual longevity and number of people living with HIV leads to an increase in the incidence of various malignancies, especially among individuals more than 50 years old, who currently comprise about 11% of ART patients in South Africa and will likely rise to 30% or more in the next 30 years (Hontelez et al., 2011). Lomalisa et al., (2000), reported that women living with HIV advance to cervical malignancy about 10 years earlier compared to women living without HIV.

HPV is a member of the *Papillomaviridae* family of viruses (Bansal et al., 2016), and known to be a causative agent of CC (Walboomers et al., 1999; Dunne and Markowitz, 2006); the second most common cancer among women worldwide (Soohoo et al., 2013; Parkin et al., 2005). The relationship between HPV and CC was exhibited in 1984 by the outstanding famous German virologist Harald zur Hausen. It is estimated that 291 million women are carriers of HPV DNA wherein 270,000 deaths annually are caused by HPV oncogenic infection, with over 85% of these occurring in developing countries (Parkin et al., 1999). Furthermore, approximately 21% of women in the general population are estimated to harbour cervical HPV at any time, and about 63% of invasive cervical cancers are attributed to HPV oncogenic types 16 or 18 (WHO, 2010).

The size of the HPV genome is approximately 7,900 bp associated with histones, while the viral particle has a diameter of approximately 55 nm (Faridi et al., 2011). To date, it has been reported that about 202 different genotypes of HPV with distinguished variations are known**.** Based on the association of HPV with CC and precursor lesions, HPV genotypes are categorised into two types: high-risk (HR) and low-risk (LR) HPV types. About 70% of CC has been reported to be caused by the most common hr (carcinogenic) HPV genotypes 16 and 18 worldwide. In contrast, lr HPV types have been reported to be non-oncogenic. Also, Khan et al., (2007) showed that types 6 and 11 are predominately involved in the development of approximately 90% of genital warts.

There are three types of prophylactic vaccines that have been developed against HPV infection globally namely: bivalent, quadrivalent and Gardasil-9. However, in South Africa only two vaccines are in use: bivalent vaccine targeting HPV type-16 and 18, quadrivalent vaccine targeting HPV type-6, 11, 16 and 18. Bivalent and quadrivalent vaccines are given to girls between the age of nine to twelve years (Moodley et al., 2013). Despite the availability of these vaccines to prevent the initial HPV infection, HPV still remains one of the leading sexually transmitted viruses (Reusser et al., 2015) making it a major public health issue that needs closer monitoring.

The Linear Array (LA) genotyping test kit (Roche) is used to screen for 37 known HPV infecting genotypes, it has the ability to detect more than one infecting genotype in the same sample. Despite its limitation to detect only 37 genotypes, LA is one of the most commonly used methodologies in the field of research. However, it is of additional value to identify HPV genotypes using more advanced molecular technologies such as Next generation sequencing (NGS) to identify infecting genotypes. NGS has the ability to read a DNA sequence millions of times giving it high specificity and sensitivity. In this regard, the technology has the ability to detect viral variants existing as a minority population in an individual. The present study investigated the presence of HPV with regards to its burden and genotype in a cohort of women in routine care in northern South Africa.

1.2 LITERATURE REVIEW

1.2.1 THE VIRAL MORPHOLOGY AND GENOMIC STRUCTURE OF HPV

The genome of HPV comprises a double stranded DNA of approximately 7,900 bp in size. It is a circular, relatively tiny, non-enveloped virus with a diameter of approximately 55 nm. It has an icosahedral capsid composed of 72 capsomers, which contain at least two capsid proteins, namely; L1 and L2. Each capsomer is a pentamer of the major capsid protein, L1 (Baker et al., 1991). Each virion capsid contains several copies (about 12 per virion) of the minor capsid protein, L2 (Sapp et al., 1995). HPV genome comprises of three major regions: early, late, and a LCR region. In all papillomaviruses, the above mentioned three regions are separated by two polyadenylation (pA) sites namely; the early (A_E) and late pA (A_L) sites. The early region of papillomavirus genomes occupies more than half of the HPV genome from its 5′ half and encodes six common open reading frames (E1, E2, E4, E5, E6 and E7) (Danos et al., 1982) that translate into individual proteins. Figure 1.1, shows the genome of HPV.

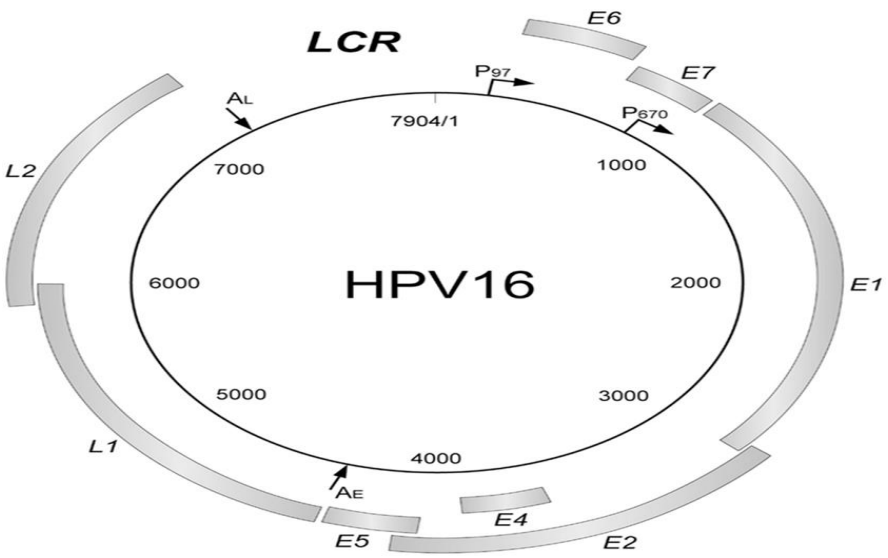

Figure 1.1: Schematic representation of the circular HPV 16 DNA genome (Adapted from Molecular biologists for Oncologists.1996).

1.2.2 LIFE CYCLE OF HPV

HPV infects cells in the basal epithelial layer. Once the virus enters the cell, the viral genome replicates inside the nucleus in a very low copy number episome (Moody and Laimins, 2010). The early genes (E1, E2, E5, E6 and E7) are then expressed in the suprabasal layer that helps in the maintenance of the HPV genome and induces the proliferation of cells, then the copies of the infected HPV cell multiplies in the epithelium, thus producing abundant copies of cells that will finally give rise to infectious virions (Hamid and Gston, 2009; Lazarczyk et al., 2009). The activation of differentiation-dependent promoter and maintenance of gene expression of the four early genes namely; E1, E2, E6 and E7 takes place in the more differentiated cells of the epithelial layer. The expression of E4 gene is then activated and its product plays a role in inducing the amplification of the replication of the viral genome, thus rapidly increasing the number of the viral copy per cell, at the same time that the expression of L1 and L2 genes takes place (Nakahara et al., 2005; Lazarczyk et al., 2009). Following, the L1 and L2 will come together to assemble the capsid and to form virions (Lazarczyk et al., 2009).

1.2.3 FUNCTIONS OF THE HPV GENES

- E1, is responsible for the replication of the virus, and is also associated with the maintenance of the genome (Raybould et al., 2011).
- E2, it regulates viral replication and transcription. The regulation can both be negative or positive, and this is accomplished by the interactions with promoters situated in the upstream regulatory region (Stern et al., 2001)
- E5, it is responsible for the down regulation of the major histocompatibility class I molecules (Maufort et al., 2010).
- E6/7, both are oncogenes responsible for the proliferation of the host cell, LCR regulates the expression levels of the mentioned oncogenes (Parfenov et al., 2014).
- E4, interacts with the cytoskeleton proteins (Raybould et al., 2011).
- L1 and L2, encoded the major and minor portions of the capsids respectively. In order to permit a proper capsid construction, the L2 region is expressed first then

L1 region after. Both genes facilitates in the viral entry into the host (Doorbar, 2005).

1.2.4 HUMAN PAPILLOMAVIRUS (HPV) CLASSIFICATION

The classification of HPV is clinically significant for various reasons: (1) only one genus of human papillomavirus is associated with CC; (2) its pathogenicity differs according to genotype. HPV is classified into 5 genera*: Alpha, Beta, Gamma, Mu and Nu-papilloma virus.* However, there are other genera that comprises papillomaviruses that are isolated from non-human mammals and birds (Drury et al., 1998; Herbst et al., 2009; Lange et al., 2011).

1.2.4.1 *ALPHA-PAPILLOMA VIRUS* (SUPERGROUP A)

This group of papillomavirus is associated with tropism for the genital epithelia. Nevertheless, common warts have been seen to be caused by some types belonging to this genus. The types found in this genus can be HR with increased risk for CC progression, example is HPV 16 and 18 classified in species 9 and 7 respectively. On the other hand, LR types such as types 6 and 11 located in species 10 are also found in this genus. In addition, the cutaneous HPVs are also found in this group (De-Villiers et al., 2004; Bernard, 2005; Hazard, 2007).

1.2.4.2 *BETA-PAPILLOMA VIRUS* (SUPERGROUP B-SUBGROUP B1)

Papillomaviruses belonging to this genus are divided into 5 various species. The most prevalent types detected in this genus belonging to species 1 are HPV 5 and 8 detected in the individuals skin with epidermodysplasia verruciformis (EV). In addition, this genus comprises of the cutaneous HPVs that can be detected in the general population without skin lesions, showing not only the ubiquity but also the high rate of non-symptomatic infections (De-Villiers et al., 2004; Bernard, 2005; Doorbar, 2005; Hazard, 2007).

1.2.4.3 *GAMMA-PAPILLOMA VIRUS* (SUPERGROUP B- SUBGROUP B2)

This genus also have 5 various species consisting of 7 various types namely; type-4, 48, 50, 60, 88, 65, and 95 (De-Villiers et al., 2004; Bernard, 2005; Hazard, 2007).

1.2.4.4 *MU-PAPILLOMA VIRUS* (SUPERGROUP E)

This group comprises of types 1 and 63. Type 1 is the cause of common and palmar warts making it the most studied member under this genus (De-Villiers et al., 2004; Bernard, 2005; Hazard, 2007).

1.2.4.5 *NU-PAPILLOMA VIRUS* (SUPERGROUP E)

HPV 41 is the only species found in this genus (De-Villiers et al., 2004; Bernard, 2005; Hazard, 2007). Figure 1.2 represents the relationships among the known various HPV genera.

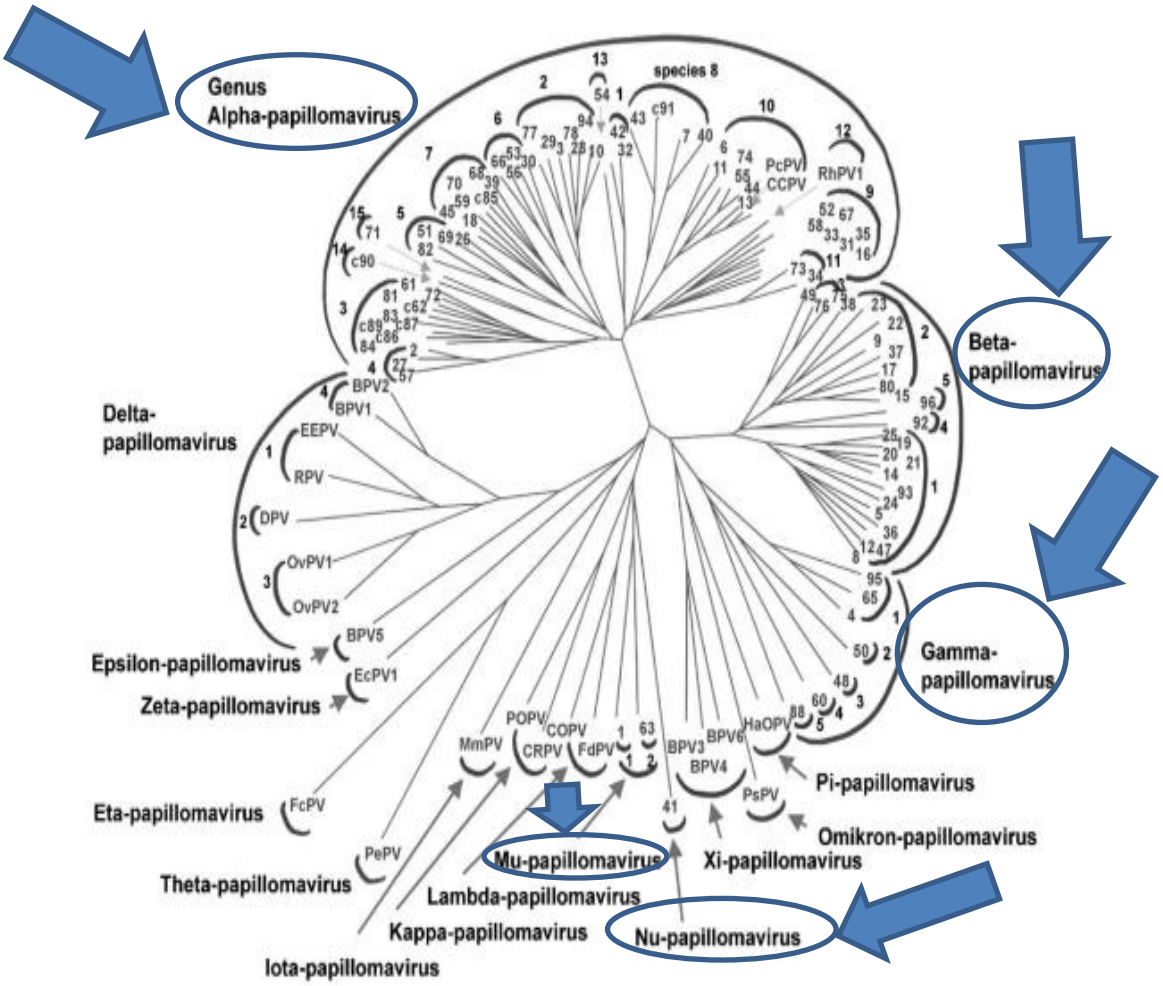

Figure 1.2: Phylogenetic tree showing various HPV types per genus (adapted from De Villiers et al., 2004).

1.2.5 RISK FACTORS ASSOCIATED WITH HPV INFECTION

A number of studies indicate that there are several risk factors associated with acquiring genital HPV and CC (Gomez and Santos, 2007). Infection by HR HPV is necessary but may not be sufficient for CC development. CC depends on the variety of additional factors (sexual factors, viral factors as well as non-viral factors) that act in concert with HPV associated types (Gomez and Santos, 2007).

1.2.5.1 VIRAL FACTORS

The progression of CC depends on the HPV type. Infection by persistent HR oncogenic HPV 16 and 18 often leads to the progression to high-grade dysplasia and invasive cancer (Ho et al., 1995). A study conducted on 1, 643 women with normal cytology, who were followed up for 4-6 years, revealed that women found with HR HPV DNA by PCR at baseline were 116 times more likely to develop CIN3 (cervical intraepithelial neoplasia) compared to women with a negative DNA test (Rozendaal et al., 1996). Compared to other HPV types, the risk of progression is approximately 40% especially in HPV types 16 and 18 (Kiviat and Koutsky, 1993).

The role of HPV variants is also an important emerging factor in the cervical neoplasia development. The variants of HPV differ chemically, biologically, as well in their pathogenicity (Giannoudis and Herrington, 2001). The geographical area and also the ethnic origin of the population play an important role in the variations of the oncogenecity of the specific variants of HPV. Based on the sequence variation of L1, L2 and URR (upstream regulatory regions of HPV-16, five variants have been defined for HPV type 16: European (E), Asian (AS), Asia-American (AA), African-1 (Af1) and African-2 (Af2). Several studies have shown that infections with multiple types of HPV can occur (Quint et al., 2001; Dickson et al., 2013).

1.2.5.2 NON-VIRAL OR BEHAVIORAL FACTORS

Several studies have shown that the major risk factor for acquiring genital HPV infection and CC is sexual activity. Having multiple sexual partners or being with someone who has multiple sexual partners increases the chances of becoming infected with HPV (Peyton et al., 2001). Factors such as high parity, overall life time

number of sexual partners, oral contraceptive usage and smoking also play a role in the acquisition of HPV (Dempsey and Mendez, 2010).

Age is also an important determinant of HPV infection. The greatest metaplastic activity occurs at puberty, first pregnancy and declines after menopause. The prevalence of HPV reaches its peak in young adults aged between 18-30 years and declines at older ages (Burk et al., 1996). In addition, it is estimated that about 46% of women in college may have HPV infection of the genital tract (Bauer et al., 1991). However, CC is more common in older women (>35 years), suggesting infection at a younger age and slow progression to cancer over time. Multiple pregnancies were a significant independent risk factor among women with histopathologic evidence of HPV infection in biopsy specimens and among women with CIN 2 and 3 (Adam et al., 2000).

1.2.6 CLINICAL COURSE AND MANIFESTATIONS OF HPV INFECTION

Infection by HPV can be categorised into three stages: dormant, subclinical or clinical stage. The viral load is an important factor in disease progression. Infection may take the pathway of low viral-load infection without clinical disease, or with clinical disease if the viral load is high (Moscicki et al., 2004). About 90% of individuals infected with low/high risk HPV clear the infection within about two years (Ho et al., 1998; Koshiol et al., 2009). It takes about six months (median time) to clear genital warts after treatment (Winer et al., 2005). Up to 30 percent of cases of genital warts especially in women spontaneously regress within four months (Lacey, 2005). In addition, women who fail to clear the infection are at risk of progression to malignancy (Moscicki et al., 2004). In women, the highest prevalence of infection by HPV takes place in their early 20s (Naucler et al., 2007). Because the peak incidence of CC does not occur until 40 years of age, testing for persistent HPV infection is most useful between 30 and 40 years of age (Naucler et al., 2007).

1.2.7 DIAGNOSIS OF HPV INFECTION

For women, the most commonly used laboratory diagnostic procedure is the pap smear. The technique is used for screening CC, whereby sampling of cells of the cervix is done and checked for signs of malignancy (Baay et al., 1996). In Pap

diagnosis, atypical squamous cell of undetermined significance (ASCUS) during cervical screening are identified. In women found with such types of abnormalities, colposcopic biopsy is collected and further analyzed (van Doorn et al., 2006). However, the molecular approaches used to diagnose HPV infection include Signal amplification and target amplification (Baay et al., 1996). Southern blot, is the gold standard for analysing HPV genome, early studies used this technique to study human papillomavirus (Southern, 1975). PCR based techniques are also used for the detection of viral DNA, It employs a series polymerization steps whereby a targeted viral genome is amplified, resulting in an increased number of HPV nucleotide sequences present in the biologic sample (Iftner and Villa, 2003).

1.2.8 THE PREVALENCE OF HPV INFECTION IN SOUTH AFRICA

Different studies document varying prevalence of HPV infection with high proportions been reported in developing countries compared to developed countries. Approximately 6000 new cases of CC are diagnosed yearly in South Africa. The World Health Organization (WHO) estimated the age-standardized incidence rate for South Africa to be 26.6 per 100 000 women (WHO/ICO, 2010). Compared to younger women, older women has been observed to have a lower HPV prevalence (De Sanjose et al., 2007; Tiggelaar et al., 2012). A study conducted by Richter et al., 2013 and Ebrahim et al., 2016 in Gauteng and KwaZulu-Natal Province (South Africa) reported a high prevalence of 85% and 76% respectively, among women less than 25 years of age. In addition, a range of 44-71% of HPV prevalence was reported in the Western-Cape Province of South Africa (Adler et al., 2014; Giuliano et al., 2015; Mbulawa et al., 2015). The HPV varying prevalence's reported in Gauteng, Kwazulu-Natal and Western Cape Province shows a higher burden of nearly 3-4 fold from the reported estimates in Sub-Saharan Africa (Richter et al., 2013), this clearly shows that geographical location plays a vital role in the distribution of HPV infection.

1.2.9 HUMAN PAPILLOMAVIRUS PROPHYLACTIC VACCINES

Vaccination against HPV infection plays a vital role in the prevention of CC. Till date, the Food and Drug Administration (FDA) had approved three vaccines against HPV infection: Bivalent offers protection against HPV-16 and 18; Gardasil, offers protection against HPV-6, 11, 16 and 18 and Gardasil-9, offers protection against

HPV-31, 33, 45, 52 and 58 plus the four HPV types covered by Gardasil. The three vaccines are designed to cover HPV types 16 and 18, the most common high risk HPV types worldwide associated with almost 70% of CC. (Gillison et al., 2008; Chaturvedi et al., 2011).

Mathematical modelling predicted the addition of virus like particles (VLPs) of HPV types 31, 33, 45, 52 and 58 to the quadrivalent vaccine could potentially protect against 90% of all cervical cancers (Serrano et al., 2012). On the other hand, HPV 6 and 11, covered by Gardasil/9 has been documented to cause about ninety present of the genital warts (Koutsky et al., 2002). Dochez et al., (2014), reported that Bivalent and Gardasil vaccines shows partial protection known as cross protection against some of the HPV types. Bivalent and Gardasil vaccines have been seen to be effective when used as prophylaxis that is administered before exposure. Since the principal route of transmission of HPV is through sexual activity, it is highly recommended that individuals get vaccinated before they engage in sexual activities.

1.2.10 NEXT GENERATION SEQUENCING

NGS is one of the most powerful technology with the ability to sequence thousands to millions of reads in a single run. To date it has gained lot of applications from various fields of research. This technology uses three main critical steps namely, library preparation, sequencing and imaging. Most of the developed commercial kits are PCR-based and targets limited or known number of HPV types that are generally found in developed countries. As a result, these commercial kits fail to pick up novel or rare HPV types that might be present in that specific population. Radford et al., (2012), reported that NGS technology has opened a lot of opportunity because of its ability to examine the viral diversity directly from the clinical specimens. In addition, Flores-Miramontes et al., (2015), showed that this technology has the ability to analyse a lot of samples at once. Variants occurring at low frequencies can also be detected by NGS technology (Ansorge, 2009).

1.3 STUDY RATIONALE

After Harold zur Hausen established the association between HPV and CC in the 1980s (Walboomers et al., 1999), it was found that infection with persistent high risk (HR) HPV type (s) is responsible for the development of intraepithelial neoplasia and

CC, the most common cancer in women worldwide (Naucler et al., 2011). Hausen, (1982) and Dillner, (2015), reported that HR HPV types accounts for almost 99% of CC cases. Although HPV infection is commonly high in women who are immuno-compromised, it is also detected at significant levels in the general population. To date, more than 202 genotypes of HPV have been identified and with the growing number of detected genotypes, Giuliano et al., (1999), Clifford et al., (2005) reported that the distribution of various HPV genotypes is influenced by various parameters such as geographical location and risk factors for acquiring HPV infection be it behavioural or viral factors. Venceslau et al., (2014) reported that of 275,000 cases of CC worldwide, about 242,000 deaths occurs in developing countries with factors such as regional differences, dispersion of major risk factors, prevalence, detection practices and availability of treatment services playing major roles (Jemal et al., 2011).

In South Africa, CC is the second most common cancer detected in women. In addition, in the year 2012, nearly 7,737 new cases of CC were reported of which 4,248 women died (South Africa human papillomavirus and related cancers, 2016). This shows that HPV infection is a major health issue that needs closer monitoring.

According to Moodley et al., (2013), South Africa has access to two vaccines: bivalent and quadrivalent vaccines that target genotypes 16 and 18 and types 6, 11, 16 and 18 respectively. Bivalent and Gardasil vaccines are preferably given to girls aged between 9-12 years. It should be noted that HPV vaccines are designed based on the known common HR infecting HPV genotypes worldwide. However, other HPV genotypes may still circulate in specific locations. Many studies conducted on HPV genotyping mainly employ commercial screening kits designed to pick up only the restricted and specific number of genotypes. Therefore, it is critical to understand the scope and the diversity of HPV genotypes in all geographic locations using both screening tools and advanced technologies such as NGS in order to validate the existing genotypes and also to increase the chances of detecting genotypes not covered by screening kits in the population, and of detecting multiple infections. This can help to identify individuals that are at increased risk for disease progression and inform efficient and effective future vaccines development.

1.4 STUDY HYPOTHESIS

The study hypothesized that there is high genetic diversity of HPV in a cohort of women in routine care in northern South Africa.

1.5 STUDY OBJECTIVES

1.5.1 MAIN OBJECTIVE

1.5.1.1 The main objective of this study was to describe the burden and genetic diversity of HPV in women in routine care in northern South Africa.

1.6 SPECIFIC OBJECTIVES

1.6.1 To determine the presence of HPV infection in a cohort of women in routine care.

1.6.2 To determine the distribution of low risk (LR) and high risk (HR) HPV genotypes.

1.6.3 To determine the distribution of single and multiple genotypic infection in the study cohort.

1.6.4 To determine the relationship between HPV positive status and associated risk factors.

CHAPTER TWO: MATERIALS AND METHODS

2.1 ETHICAL CONSIDERATIONS

The research protocol of this study was approved by the Research Ethics Committee of the University of Venda, South Africa (SMNS/17/MBY/07). Permission to access public Health care facilities and collection of the cervical specimen was obtained from the Department of Health, Limpopo Province. Prior to initiation of collection of demographic information, clinical data and cervical specimens, a signed informed consent was obtained from each study participant with the help of qualified health care practitioners. Furthermore, to avoid breach of confidentiality, the names of study participants were replaced with codes in sampling, sample processing, and data analysis.

2.2 STUDY SITES AND STUDY POPULATION

Study participants were recruited from the Rethabile Community Health Centre, Seshego Health Centre, Thohoyandou Health Centre, and Donald Fraser Hospital located in the Limpopo Province in northern South Africa. These are public health care facilities. Cervical specimens were collected between March and August 2017 from a total of 87 women with unknown HPV infection history. In addition, cervical specimens were collected from both HIV positive and negative women. Figure 2.1 represents the location of the health care facilities in northern South Africa from where study participants were recruited.

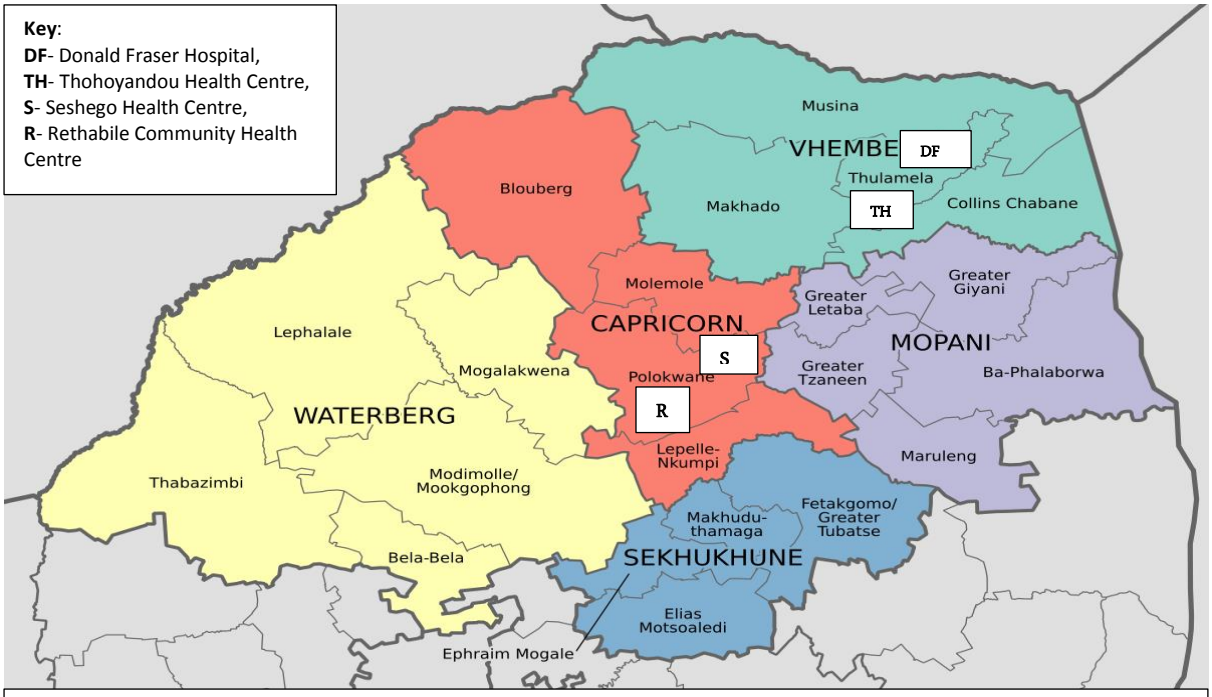

Figure 2.1: Schematic representation of the study sites, adapted from **https://en.wikipedia.org/w/index.php?title=List_of_municipalities_in_Limpopo&oldid=860431908.**

2.3 INCLUSION AND EXCLUSION CRITERIA OF RECRUITING STUDY PARTICIPANTS

The eligibility of participants to be enrolled in this research study was determined by the following inclusion criterion: women referred by a Physician for Pap smear testing, and non-referrals (women requesting Pap smear as part of their routine care) were recruited. Individuals excluded from the study were women who were on their menstrual cycle or having any vaginal discharge, pregnant, those who were on treatment for any sexually transmitted infection, non-permanent residents of Limpopo Province, and those women having a history of hysterectomy. All participants received health education talks related to the study from the health officials prior to demographic, clinical and specimen collection.

2.4 PROTOCOL USED DURING COLLECTION OF CERVICAL SPECIMENS

Cervical specimens were strictly collected by a qualified healthcare practitioner in the Pap-smear room using Aptima Cervical Specimen Collection and Transport Kit (Hologic, San Diego, CA). The kit contained each of the following; cervical brush

(collection device), cleaning swab and a tube containing Aptima Specimen Transport Medium (4.3 mL).

For each study participant, a once off sterile plastic vaginal speculum (Orthopaedic textiles, Pty, South Africa) was used to ensure the full view of the cervix. In the case where a participant had excess mucus from the cervix, a cleaning swab was used to remove the mucus then discarded into a bio-hazardous bin. The cervical specimen collection device (brush) was inserted into the endocervical canal; to ensure adequate sampling the collection device was rotated clockwise three full turns. The cervical brush was carefully withdrawn from the cervix, and cervical cells were immediately dislodged in the transport medium by rotating the collection device three times making sure that the tube was tightly re-capped. Specimens were stored at 4˚C and transported to the laboratory at the University of Venda at most within 48 hours for processing.

2.5 PROCESSING OF CERVICAL SPECIMEN AND DNA EXTRACTION

The cervical specimen was transferred into a 10 ml sterile conical centrifuge tube and centrifuged for 10 minutes at 3000 revolution per minute (rpm). After centrifugation, the supernatant was immediately aspirated into a new clean 10 ml tube and stored at -80^0 C freezer for future use. The pellet was used for DNA extraction.

Total DNA was extracted from the resulting cell pellet retained, after centrifugation, using QIAamp DNA mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Prior to the extraction procedure, the cell pellet was resuspended in 200 µl of phosphate-buffered saline (PBS X1, pH7.4). Extracted DNA was stored at -20˚C until used.

2.6 ROCHE LINEAR ARRAY GENOTYPING TEST KIT

2.6.1 AMPLIFICATION OF HUMAN PAPILLOMA VIRUS DNA (L1 REGION, 450 bp)

Polymerase chain reaction (PCR) is a powerful scientific technique used for the amplification of single or low copies of nucleic acids resulting into thousands to millions copies of the original DNA sequence (Joshi and Deshpande, 2010). This technique was developed by an American biochemist called Kary Mullis in 1983 (Bartlett and Stirling, 2003). To date, it has gained a variety of applications from

clinical, medical and research purposes worldwide. For this study, one round conventional PCR was performed to amplify a 450 bp fragment from the HPV L1 capsid region to determine the presence of various low and high risk HPV genotypes using Roche linear array HPV genotyping test kit (Roche Molecular Systems) that comes with its own HPV master mix and controls (+/-) respectively. Following the manufacturer's protocol, 50 µl LA-HPV master mix and 50 µl of the extracted DNA was achieved into a final reaction volume of 100 µl. Immediately after amplification, 50 µl of the PCR amplicon was aliqouted into two separate 2 ml sterile Eppendorf tubes. This was done to achieve genotyping by both LA and NGS methods. For LA genotyping, the aliquoted PCR tubes containing the amplicons each was added with LA denaturation reagent and gently mixed by pipetting the solution up and down five times. The other 50 µl of the PCR amplicons were stored in 4˚C and subsequently used for NGS analysis. Table 2.1 shows the detailed cycling parameters that were used during amplification of HPV DNA.

Table 2.1: Cycling conditions used during the amplification of HPV DNA (Roche Molecular Systems, Inc. Branchburg, NJ USA).

Cycling conditions	**Temperature (^{0}C)**	**Time (seconds,sec or minutes, min)**
Hot start	50	2 min
	95	9 min
Amplification 40 X (cycles)		
Denaturation	95	30 sec
Annealing	55	1 sec
Elongation	72	1 min
Final elongation	72	5 min
Holding at 72°C for not more than 4 hours prior to denaturation of the PCR amplicons (according to the manufacturer's protocol)		

Ramp rate= 50%

2.6.2 VISUALIZATION OF PCR OUTCOMES IN A 1% AGAROSE GEL

Gel electrophoresis was used to resolve all PCR products. Amplicons were resolved at 300 Amperes and 80 volts (V) for 35 minutes. One percent (1%) agarose gel stained with 0.5 ug/ml of ethidium bromide was used. From the undenatured PCR amplicons 3 µl of each PCR amplicon was mixed with 1 µl of DNA loading dye and loaded into each well to confirm which specimens are positive for HPV DNA. A

100bp molecular marker was used for size confirmation; agarose gel was viewed under UV trans-illumination gel documentation system (Syngene G, Germany). All positive specimens by PCR with expected band size of 450 bp were subjected to hybridization followed by genotyping using the Linear Array Detection Kit (Roche) (using the denatured PCR amplicons explained previously in section 2.7). Note, the remaining 47 µl was stored in -20^0 C was used for Next Generation sequencing.

2.6.3 QUALITY CONTROL MEASURES DURING PCR

Proper collection, handling and transportation of samples play a major role in getting good results during downstream application(s). Contamination of PCR experiments remain one of the major problem especially in Microbiology laboratories; and it appears that the sources of contaminants are diverse and arise from different sources (reagents, disposables, water, work environment and amplicons). Hence to avoid getting false positive results, freshly prepared reagents, sterile tubes and water should always be used at all times.

The negative and positive controls for all PCR reactions are used to confirm if the protocol was working as expected and to check if there was no form of contamination during PCR. In addition, a separate dedicated area and equipment was maintained for activities such as preparation of samples, setup of PCR and handling of amplified products. The PCR work area was regularly decontaminated using 10% bleach, 70% ethanol and by ultraviolet (UV) irradiation.

2.6.4 HYBRIDIZATION AND GENOTYPING OF HPV DNA (USING DENATURED PCR AMPLICONS)

Following denaturation of the PCR amplicons using LA reagents, the amplicons were subjected to hybridization using linear array detection kit (Roche Molecular Systems, Inc. Branchburg, NJ USA). Using a heat resistant permanent marker, the genotyping strips were properly labelled with the specimen codes and controls identification. Thereafter, 75 µl of the denatured PCR amplicons was hybridized and used to detect the HPV genotypes in a 24-well tray lid. The HPV reference guide was used manually to interpret the LA HPV genotyping strips.

2.7 NEXT GENERATION SEQUENCING

2.7.1 PURIFICATION OF THE UNDENATURED POSITIVE AMPLICONS BY PCR

The undenatured portion of HPV DNA for each sample was purified using AMPure XP beads protocol. The beads were warmed at room temperature for 30 minutes before use followed by brief vortexing. Forty microliters and thirty microliters of the PCR amplicons and AMPure XP beads respectively were added to a 2 ml Eppendorf tube. The solution was well mixed by brief vortexing, and centrifuged for 2 minutes at 1800 rpm followed by 5 minutes incubation at room temperature. Thereafter, the tube containing the mixture was placed on a magnetic stand for approximately 2 minutes until the liquid cleared, the supernatant was then discarded. A two sequential washing step was performed using freshly prepared 80% molecular grade ethanol. The beads were re-suspended in a volume of 52.5 µl RBS (re-suspension buffer), all these was done following the manufacturer's recommendations.

2.7.2 QUANTIFICATION OF THE PCR AMPLICONS

The purified amplicons were quantified using dsDNA high sensitivity assay kit and Qubit 3.0 Fluorometer. Before quantification, all the kit reagents were warmed at room temperature. For each assay, two standards 1 and 2 were set up with the test samples into a final volume of 200 µl separately and prepared as follows: 190 µl of the Qubit working solution was mixed with 10 µl of standard 1 and 2 separately inside the Qubit assay tube. In addition, for the test samples: 198 µl of the Qubit working solution was mixed with 2 µl of the purified samples into a Qubit assay tube, all the tubes were vortexed for approximately 3 minutes, and then followed by 2 minutes incubation at room temperature, the tubes were read individually following Qubit 3.0 Fluorometer instructions.

2.7.3 LIBRARY PREPARATION AND SEQENCING

All the quantified double stranded DNAs were diluted into a final concentration of 1ng. The normalized genomic DNA was then used as an initial input for tagmentation. To fragment the individual amplicons, Nextera kit comprising of the transposon technology was used to cut the DNA randomly and add sequencing adapter in the process. Following tagmentation, amplification of libraries was performed in order to add the molecular tags; and to remove all the unadded tags

the libraries were cleaned up using AMPure XP beads following the manufacturer's protocol, and then quantified using Qubit ds high sensitivity kit. To verify the average size of the fragmented amplicons, amplicons were subjected to a 1% E-gel. Depending on the library size, all the samples were manually pooled at an equimolar ratio of 1 nM per sample. The pooled libraries were denatured with 0.1 N NaOH and diluted to a final volume of 500 µl at 1.8 pM, following the Illumina MiniSeq instructions. For a control, 20% of the denatured PhiX genome was then added, to the dilution. The Libraries were then loaded into MiniSeq High Output cartridge 300 cycles and sequenced using the Local Run Manager option.

2.7.4 SEQUENCE ANALYSIS

Generated raw sequence data was downloaded from local run manager as fastaQ format. The generated data was then subjected to fastQC to check the quality of the reads before analysis. Data was imported into Genious software version 11.0.5 wherein the forward and reverse reads were paired. To clean the incorrect base callings, the 3' and 5' ends were trimmed using the Error Probability Limit of 0.001 found under annotate and predict option. Furthermore, the reads were then mapped with all HPV reference sequences belonging to L1 Alpha species downloaded from Papillomavirus Episteme (PaVE) database. This was employed to generate consensus sequences that were used for phylogenetical analysis. It is important to mention that only the mapped sequences were used for phylogenetic analysis, references sequences showing no contigs found were considered unmapped then excluded from the analysis. Genious tree builder (Phylogenetics) was then employed to study the evolutionary association of the test samples and the references. Moreover, to compute a tree, Neighbour-Joining method was selected. A general tree comprising all the L1 Alpha species reference sequences for each participant was drawn to confirm the genotypes, thereafter only the clustered references with the test samples were then used to compute a final tree for this current study.

2.8 STATISTICAL ANALYSIS

To determine the relationship between positive HPV status and associated risk factors, participant's demographic and clinical data (only positive specimens by PCR) was analysed using SPSS software package version 22.0. A univariate

analysis was performed to compare the constant variable (HPV status) with the independent variables (risk factors). The Chi-square test was performed to determine the significance of 11 risk factors namely: age, smoking status, highest level of education, occupation status, age at first sexual intercourse, marital status, HPV vaccination status, gender of a sexual partner, Pap smear testing status, HIV status and number of sexual partners in relation to HPV positive status. A p-value of less than 0.05 was used as a cut-off for significance, a p-value less than or equal to 0.05 was considered statistically significant.

CHAPTER 3: RESULTS

3.1 DEMOGRAPHIC AND CLINICAL PROFILE OF THE STUDY COHORT

A total of 87 women participated in this study however; complete demographic and clinical data was available for 81 participants. The mean age of the women was 40 years, with a range of 24-67 years. Married, divorced and widows were in the following decreasing order of frequency: n= 31; 38%; n=5; 6%; and n=3;4% respectively, about half of the women (52%, n=42) were single. Eighty percent (n=65) of the women had 1 sexual partner by the time this current study was conducted. Majority of the women (n=71; 87.7%) reported they have never been vaccinated against HPV infection. Among the respondents, the lowest age at first sexual intercourse was 13 years. Furthermore, 38% (n=31) of the women reported doing their first ever Pap smear at the time of the study. Fifty six reported to be HIV infected (69.1%), and 25 reported to be HIV negative (30.8%). Some participants were not comfortable disclosing all the required personal information. The detailed demographic profile of the study participants is showed in table 3.1.

Table 3.1: Demographic and clinical characteristics of the study population (n=81)

Characteristics	Distribution (%)
Gender	
Women	81 (100%)
Age (years)	
Mean	40
Range	24-67
Marital status	
Single	42 (52%)
Married	31 (38%)
Divorced	5 (6%)
Widow	3 (4%)
Highest level of education	
Never attended school	1 (1.2%)
Grade 1-7	8 (9.9%)
Grade 8-12	55 (67.9%)
Varsity/College	17 (21%)
Occupation (81)	
Employed and unemployed	40 (49.4%) each
Pensioner(s)	1 (1.2%)
HIV status (81)	
Positive	56 (69.1%)
Negative	25 (30.8%)
Smoking status (81)	
Smokers	8 (9.9%)
Non-smokers	72 (88%)
No response	1 (1.2%)
Age at first sexual intercourse (81)	
Range	13-35
13-18	42 (51.9%)
19-24	27 (33.3%)
≥25	4 (4.9%)
No response	8 (9.9%)
Number of sexual partner(s) 81	
0	2 (2.5%)
1	65 (80.2%)
2	3 (3.7%)
Celibate	2 (2.5%)
No response	9 (11.1%)
HPV vaccination status (81)	
Never been vaccinated	71 (87.7%)
No response	10 (12.3%)
Pap smear testing status (81)	
1st time	31 (38.3%)
2nd time	20 (24.7%)
3rd time	1 (1.2%)
Three times a year	9 (11.1%)
Yearly	12 (14.8%)
No response	8 (9.9%)

3.2 AMPLIFICATION AND SCREENING OF HPV (L1 GENE 450bp) BY LINEAR ARRAY (PCR)

In this study, 87 DNA cervical specimens (100%) were subjected to PCR in order to amplify the L1 fragment of HPV comprising of 450 base pairs (bp). Following PCR, the amplicons were viewed under 1% agarose gel to confirm the presence of HPV infection in the study cohort with expected band size.

Of the 87 participants, 20 (23%) were positive for HPV DNA by PCR. Of the 20, 16 (80%) were HIV positive (p-value 0.077). Figure 3.1, is a representative of a gel picture designating some of the positive samples by PCR.

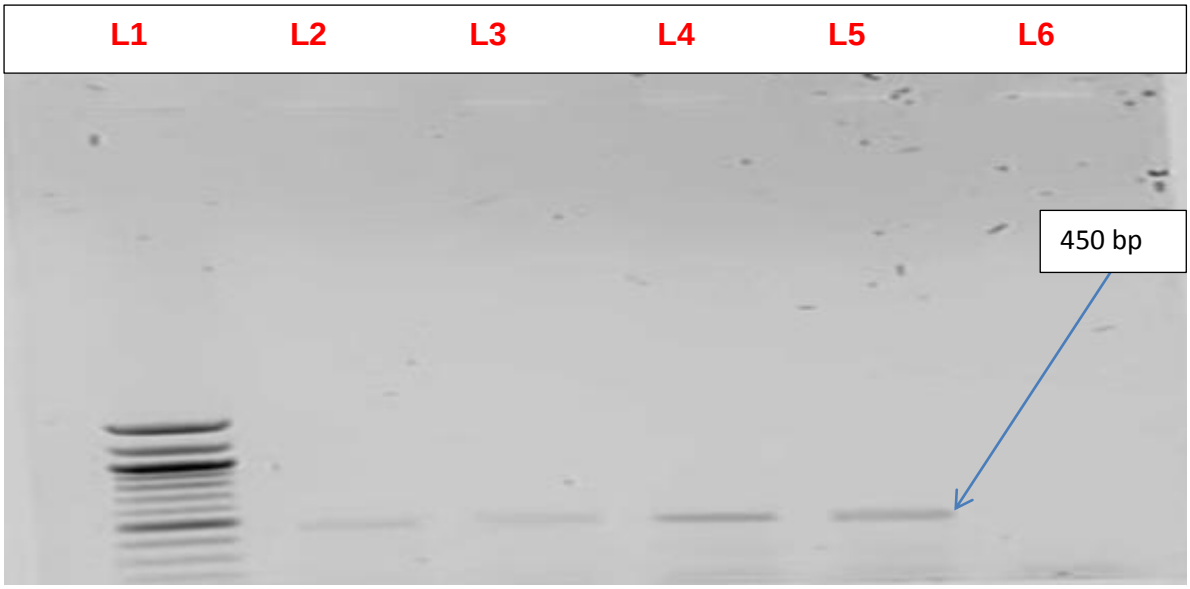

Figure 3.1: A representative of 1% agarose gel: Lane 1 (L1) is a 100 bp molecular weight marker; L2-L5 are positive test samples and L6 is a negative control respectively.

3.3 HPV GENOTYPES DETECTED BY LINEAR ARRAY HYBRIDIZATON METHOD

Genotyping data was successfully generated for all the 20 positive amplicons by PCR. Figure 3.2 shows the results that were obtained after hybridization while figure 3.3 illustrates the LA reference guide that was used to identify the existing genotypes in the study cohort.

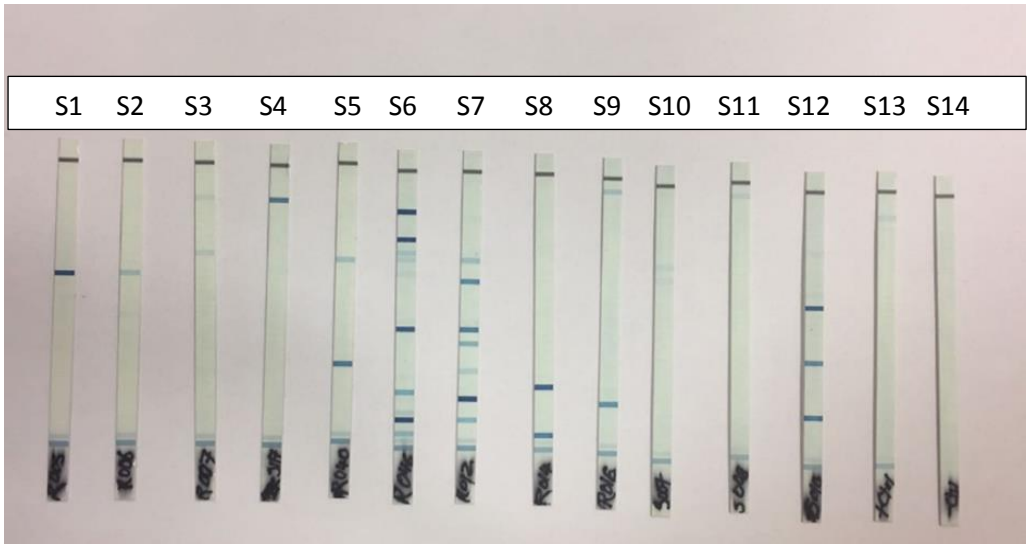

Figure 3.2: Shows the genotyping strips of the test specimens that were successfully hybridized and were ready for genotyping. Genotyping strips, S1-12 are test specimens, 13 and 14 is a positive and negative control respectively. The first band on top of each strip is a reference, the last two bands on the bottom are beta low and high globins that serves as controls too. Lastly the blue bands in between the reference band and the beta globins are the detected genotypes per participant.

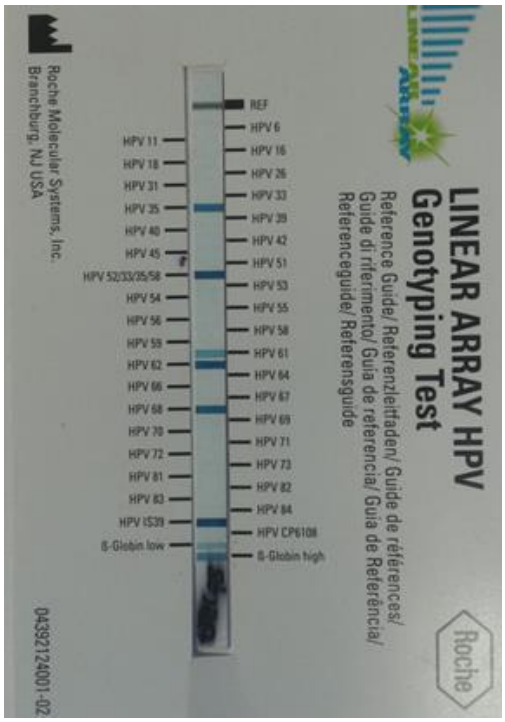

Figure 3.3: LA genotyping reference guide was used to identify the genotypes that were present in the study population. Specimen code ORHC-043 above is used as an example to show how each of the genotypes on the hybridized strips were detected. The detected genotype on this specimen included HPV types 35, 52, 61, 62 and IS39.

Overall, 27 (73%) of the 37 HPV genotypes incorporated in the Roche Linear Array method were detected. The detected genotypes include: types 84, 83, 81, 73, 72, 71, 70, 69, 68, 66, 62, 61, 59, 54, 53, 52, 51, 45, 42, 39, 35, 26, 18, 16, 6, IS39 and CP6108. The ten HPV genotypes incorporated in the Roche Linear Array method which were not detected in the study population included: types 82, 67, 64, 58, 56, 55, 40, 33, 31, and 11.

3.4 DISTRIBUTION OF SINGLE AND MULTIPLE INFECTIONS

Fifteen of the 20 women (75%) in whom HPV was detected by PCR had multiple HPV infections. Table 3.2: Shows the distribution of single and multiple infections among the 20 women infected with HPV.

Table 3.2: The distribution of single and multiple infections in participants in whom HPV was detected.

Study subjects	Genotypes detected	HIV status
1. AHDR-S117	❖ HPV type 18	Positive
2. AHDR-R297	❖ HPV types16, 62, 68 and 81	Positive
3. AHDR-R300	❖ HPV types 18, 59 and 73	Positive
4. AHDR-R308	❖ HPV type 66	Positive
5. ORHC-005	❖ HPV type 53	Positive
6. ORHC-006	❖ HPV type 53 and 61	Positive
7. ORHC-007	❖ HPV type 18 and 45	Positive
8. ORHC-012	❖ HPV types 45, 70, 53, 81, 62, 66 and 84	Positive
9. ORHC-014	❖ HPV type 72 and CP6108	Positive
10. ORHC-016	HPV types 16, 72 and 81	Positive
11. ORHC-028	❖ HPV type 83 and CP6108	Positive
12. ORHC-031	❖ HPV types 39, 69, 70 and 81	Positive
13. ORHC-032	❖ HPV types 59, 62 and 83	Positive
14. ORHC-040	❖ HPV type 51 and 70	Positive
15. ORHC-043	❖ HPV types 35, IS39, 61, 62, 68 and 52	Positive
16. ORHC-046	❖ HPV types 84, 83, 81, 73, 71, 62, 61, 52, 51, 45, 42, 39, 26 and CP6108	Positive
17. OSHC-007	❖ HPV type 42 and 51	Negative
18. OSHC-008	❖ HPV type 6	Negative
19. OSHC-010	❖ HPV types 81, 66, 52, 54 and 35	Negative
20. OTHC-014	❖ HPV type 16	Negative

Sixteen of the twenty women were HIV positive. The distribution of HPV infection among the study subjects was as follows: 25% had single infection, 30% and 15% of participant's harboured double and triple infections respectively. Meanwhile those who were seen to be having more than 3 HPV types were 30%.

3.5 DISTRIBUTION OF LR HPV INFECTING GENOTYPES

In this study, thirteen LR HPV genotypes namely: types 84, 83, 81, 72, 71, 70, 62, 61, 54, 42, 6, IS39 and CP6108 were detected by LA. This distribution was observed

in 14 (70%) HPV positive women of which 11 of them were living with HIV. The most frequent LR HPV genotypes observed in order of decreasing frequency were HPV: types 81; n=6 (43%), 62; n=5 (36%); 61 and 83 each with n=3 (21%); HPV types 84, 72, 70, 42 and CP6108 with each having a frequency of 14% (n=2). The least HPV genotypes detected under LR types with n=1 was HPV type; 71, 54, 6 and IS39 each with a frequency of 7%.

3.6 DISTRIBUTION OF HR HPV INFECTING GENOTYPES

Of the 20 HPV positive women, 17 (85%) women harboured at least 1 of the following 14 HR HPV genotypes namely: types 73, 69, 68, 66, 59, 53, 52, 51, 45, 39, 35, 26, 18 and 16. The most common genotypes observed in order of decreasing frequency under HR was HPV types: 66, 53, 52, 51, 45, 18 and 16 harboured by n=3 (18%) respectively, n=2 (12%) harboured HPV type: 73, 68, 59, 39 and 35 respectively. The least detected genotype under HR with n=1 (6%) was HPV type 26 and 69.

3.7 NEXT GENERATION SEQUENCING RESULTS

3.7.1 HPV GENOTYPES DETECTED BY NGS

The other portion of the undenatured positive amplicons by LA PCR was subjected to Next Generation Sequencing (NGS). Next Generation sequencing was employed for validating the detected HPV genotypes by LA and for further identification of the genetic variants of HPV types not picked up by the LA genotyping assay.

After sequencing, genotyping data was successfully generated from 15 of 20 (75%) positive amplicons by PCR. Following from figure 3.4-3.18 are phylogenetic trees of all the 15 successful specimens that was sequenced, it important to know that only detected genotypes were incorporated on the trees.

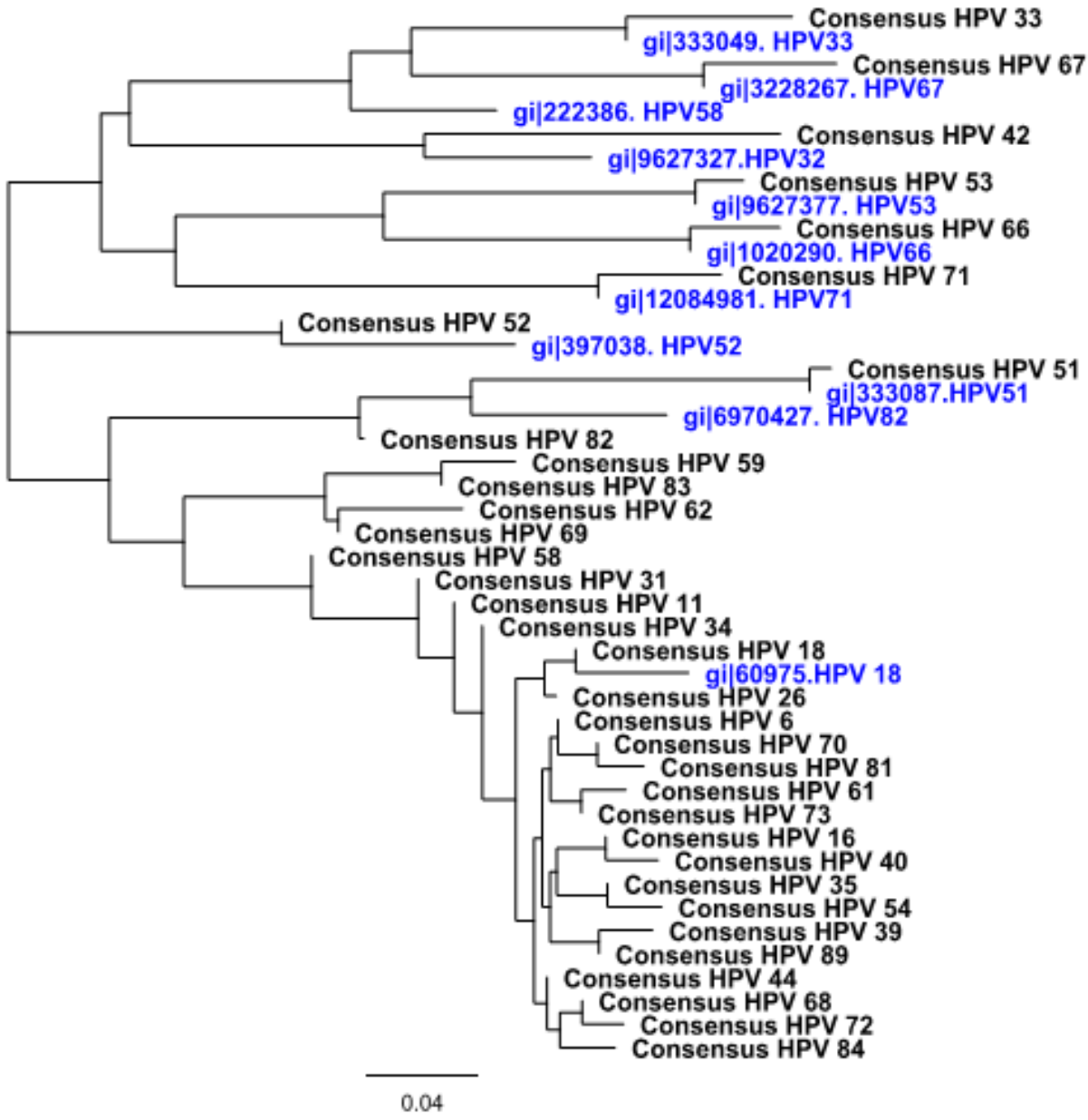

Figure 3.4: A representative phylogenetic tree of sample identity AHDR-S117. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping AHDR-S117 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 82, 71, 67, 66, 58, 53, 52, 51, 33, 32 and 18 for sample identity AHDR-S117, it is important to note that only the detected genotypes were used to compute this tree.

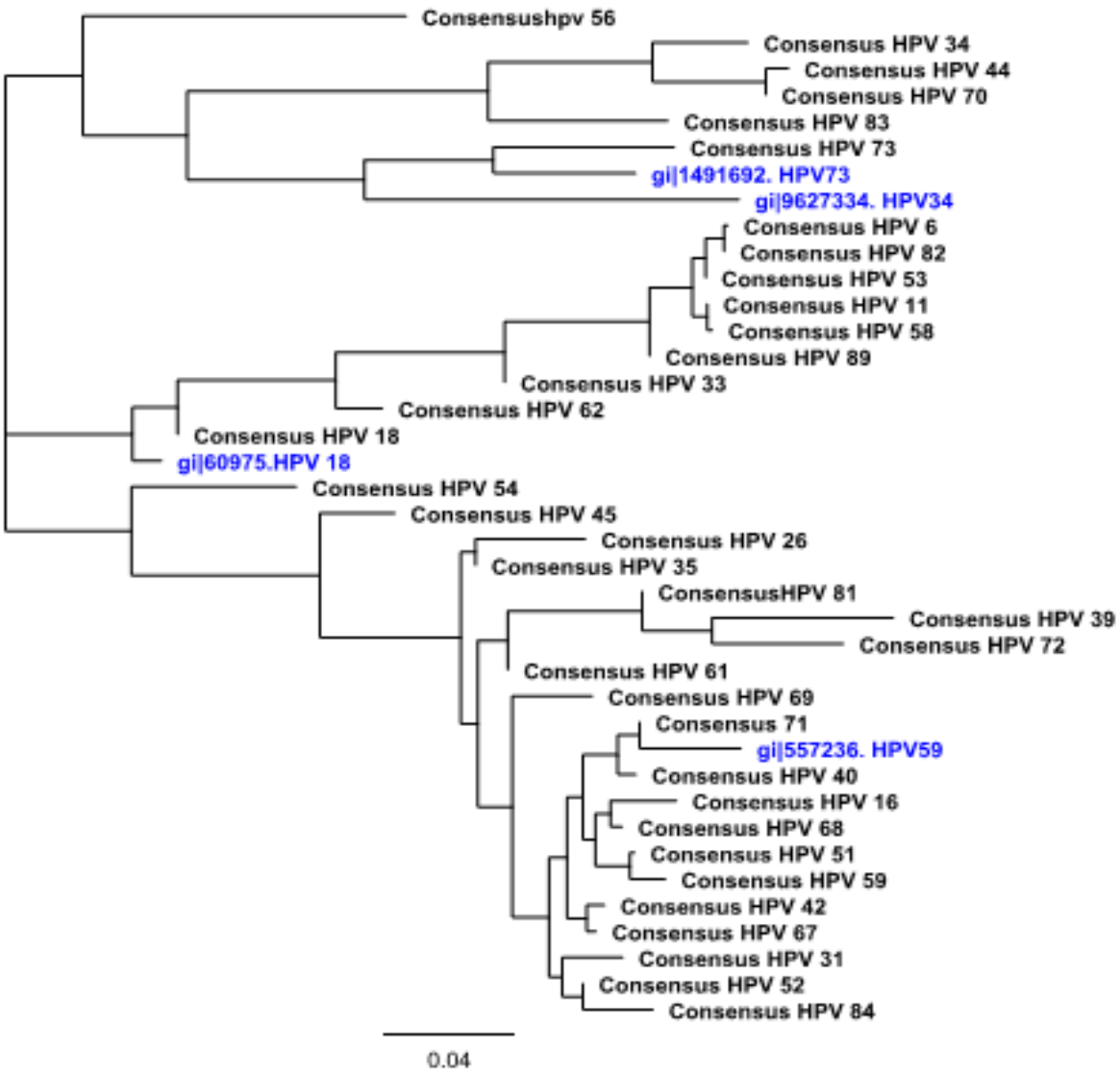

Figure 3.5: A representative phylogenetic tree of sample identity AHDR-R300. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping AHDR-R300 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 73, 59, 34 and 18 for sample identity AHDR-R300, it is important to note that only the detected genotypes were used to compute this tree.

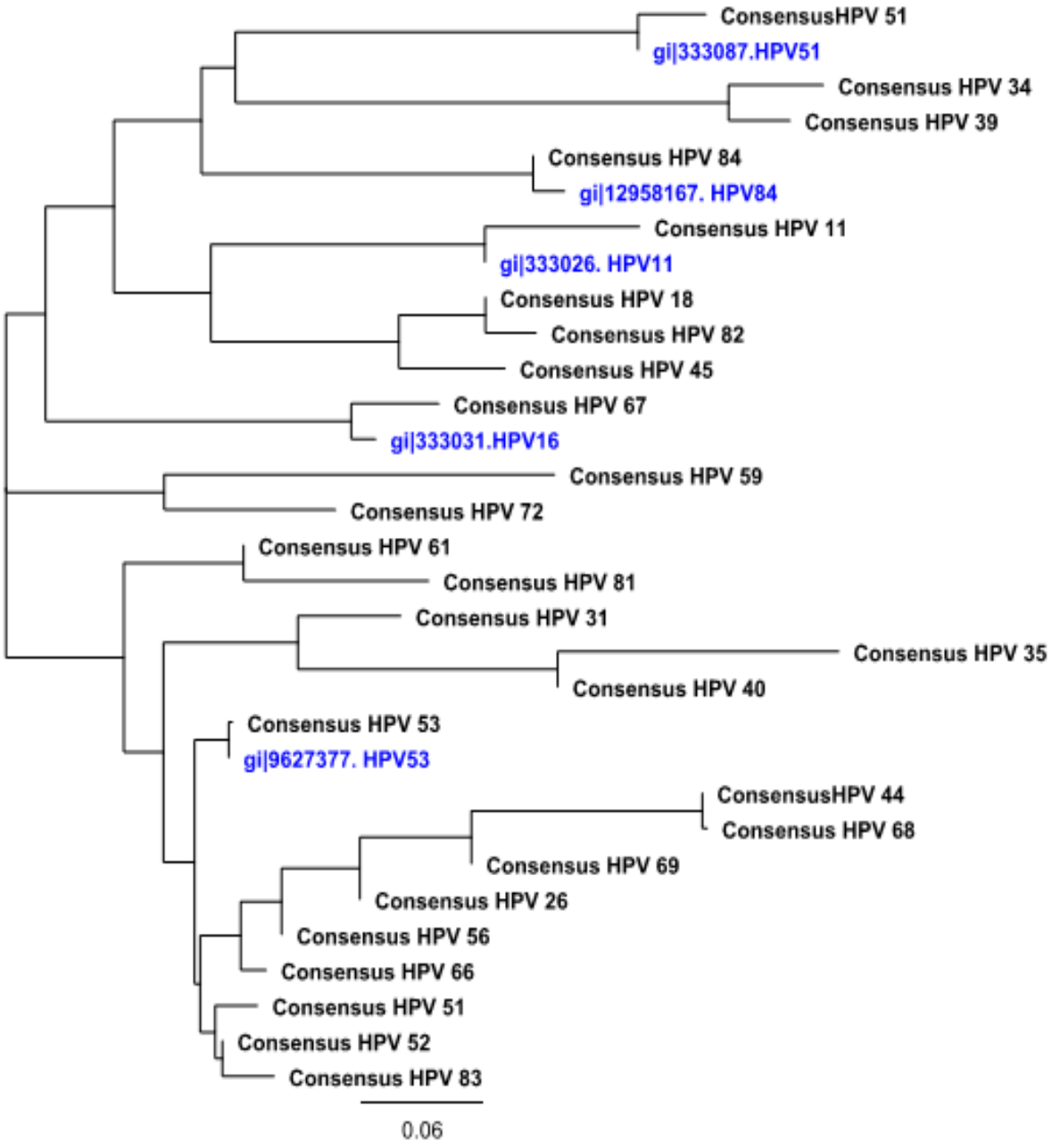

Figure 3.6: A representative phylogenetic tree of sample identity ORHC-005. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-005 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV 84, 53, 51, 16 and 11 for sample identity ORHC-005, it is important to note that only the detected genotypes were used to compute this tree.

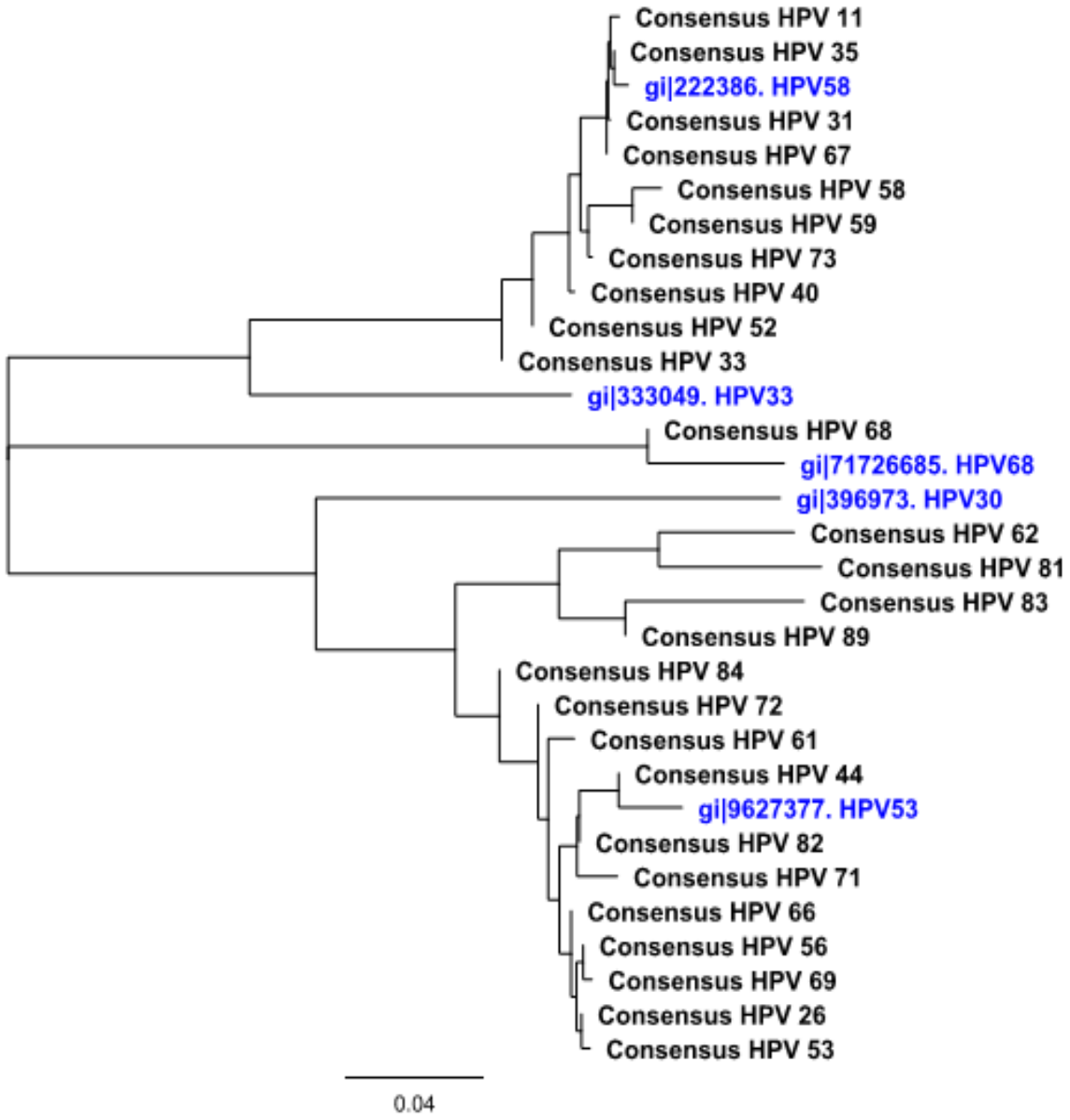

Figure 3.7: A representative phylogenetic tree of sample identity ORHC-006. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-006 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 68, 59, 53, 33 and 30 for sample identity ORHC-006, it is important to note that only the detected genotypes were used to compute this tree.

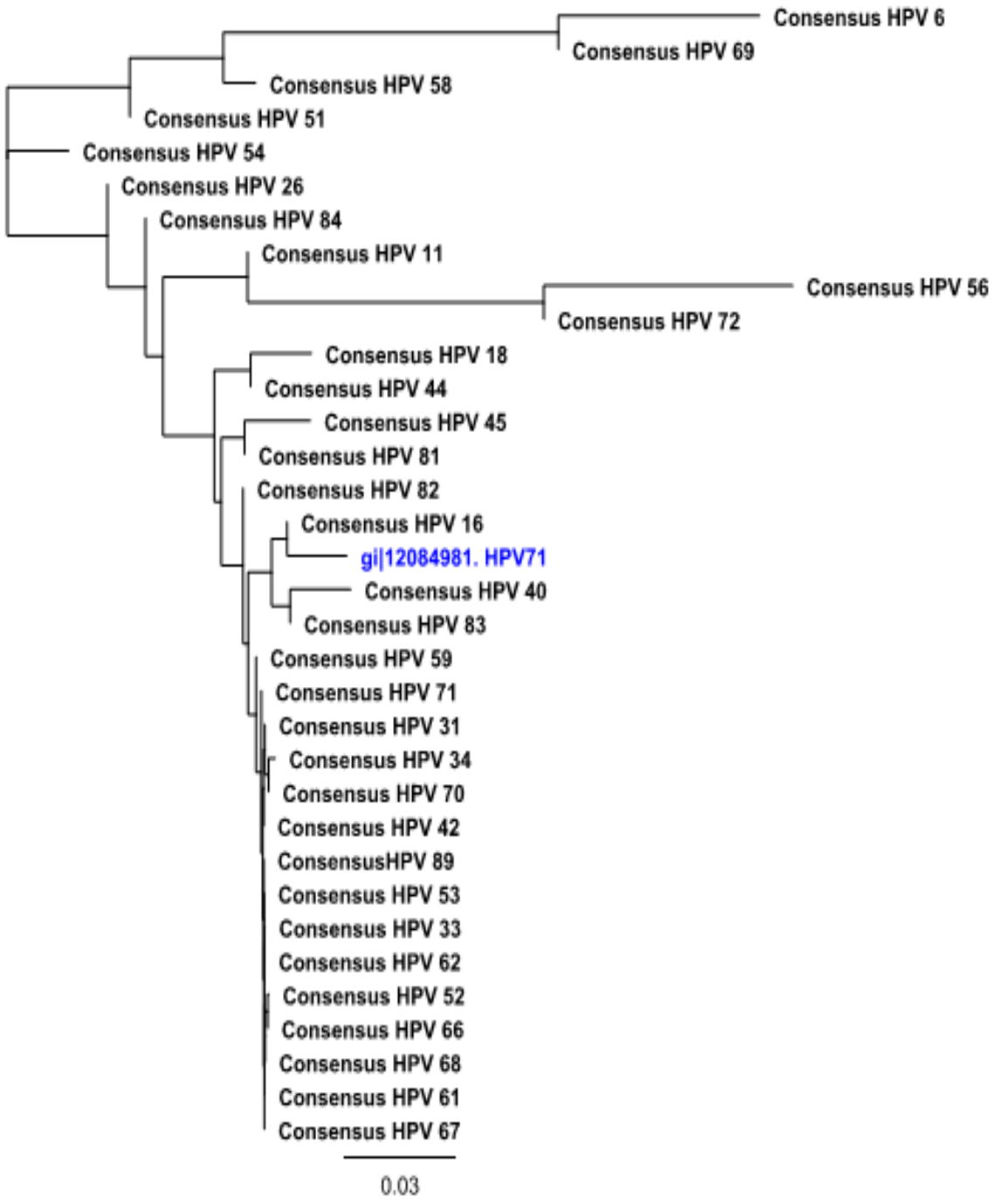

Figure 3.8: A representative phylogenetic tree of sample identity ORHC-007. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-007 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequence in blue illustrates the detected genotype namely HPV type 71 for sample identity ORHC-007, it is important to note that only the detected genotypes were used to compute this tree.

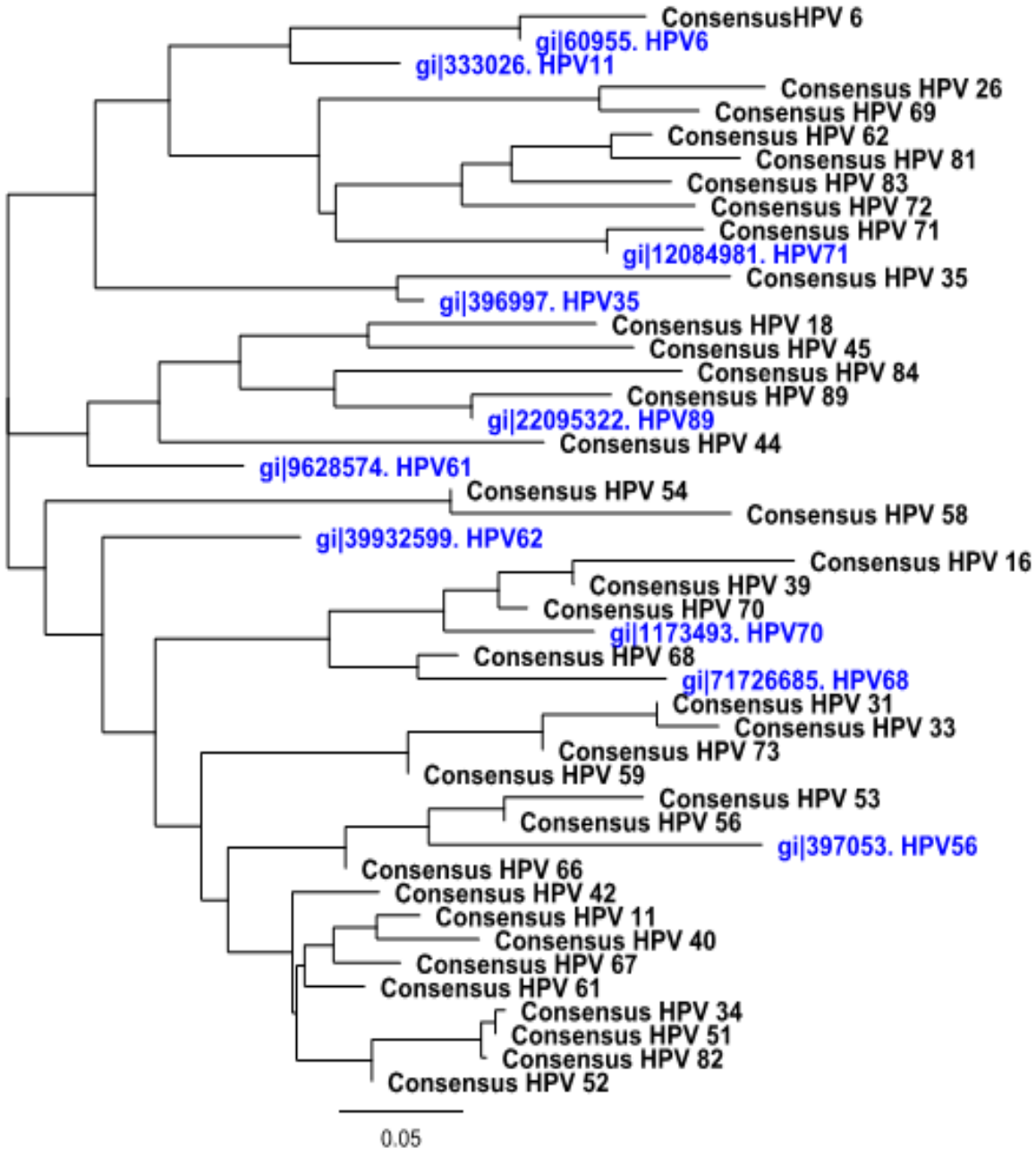

Figure 3.9: A representative phylogenetic tree of sample identity ORHC-012. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-012 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 89 (CP6108), 71, 70, 68, 62, 61, 56, 35, 11 and 6 for sample identity ORHC-012, it is important to note that only the detected genotypes were used to compute this tree.

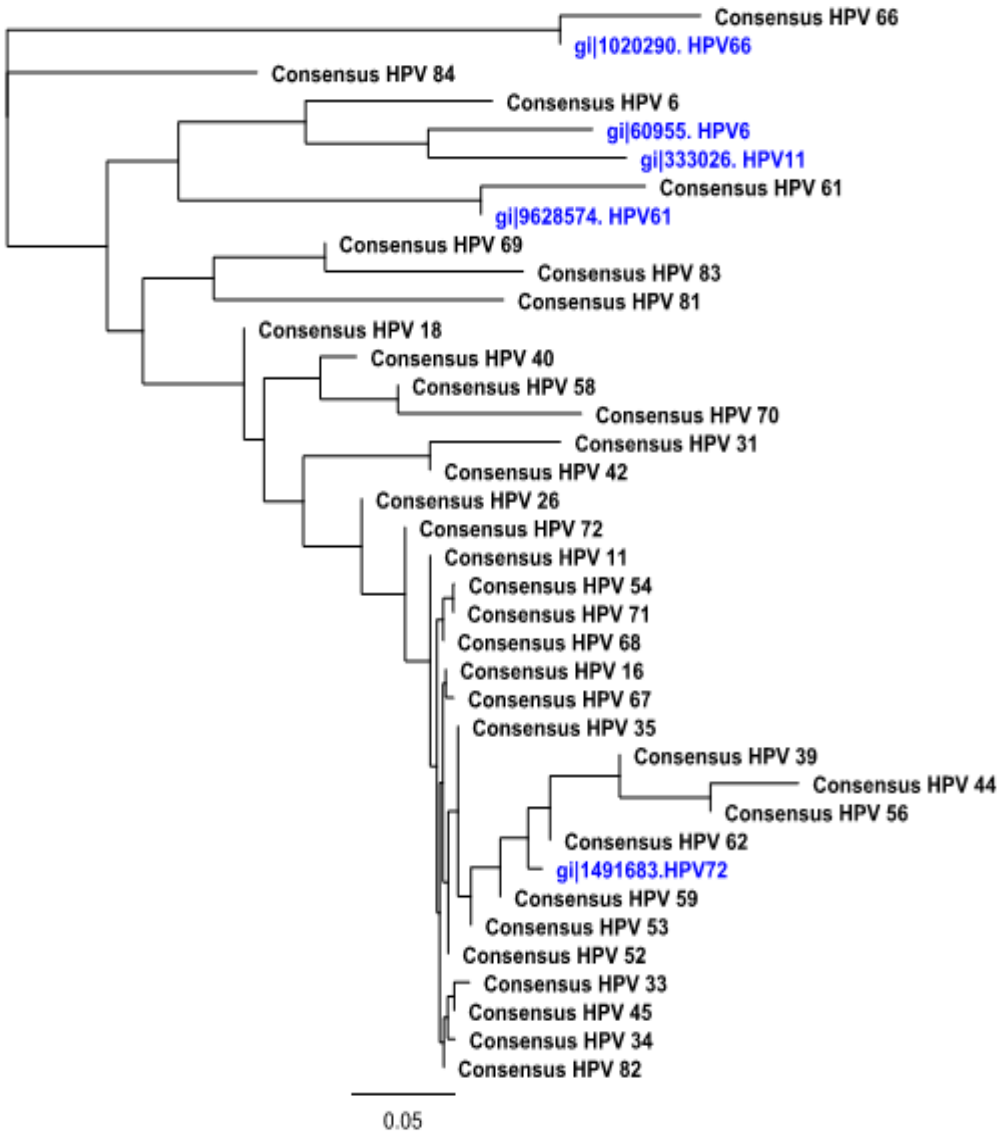

Figure 3.10: A representative phylogenetic tree of sample identity ORHC-014. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-014 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 72, 66, 61, 11 and 6 for sample identity ORHC-014, it is important to know that only the detected genotypes were used to compute this tree.

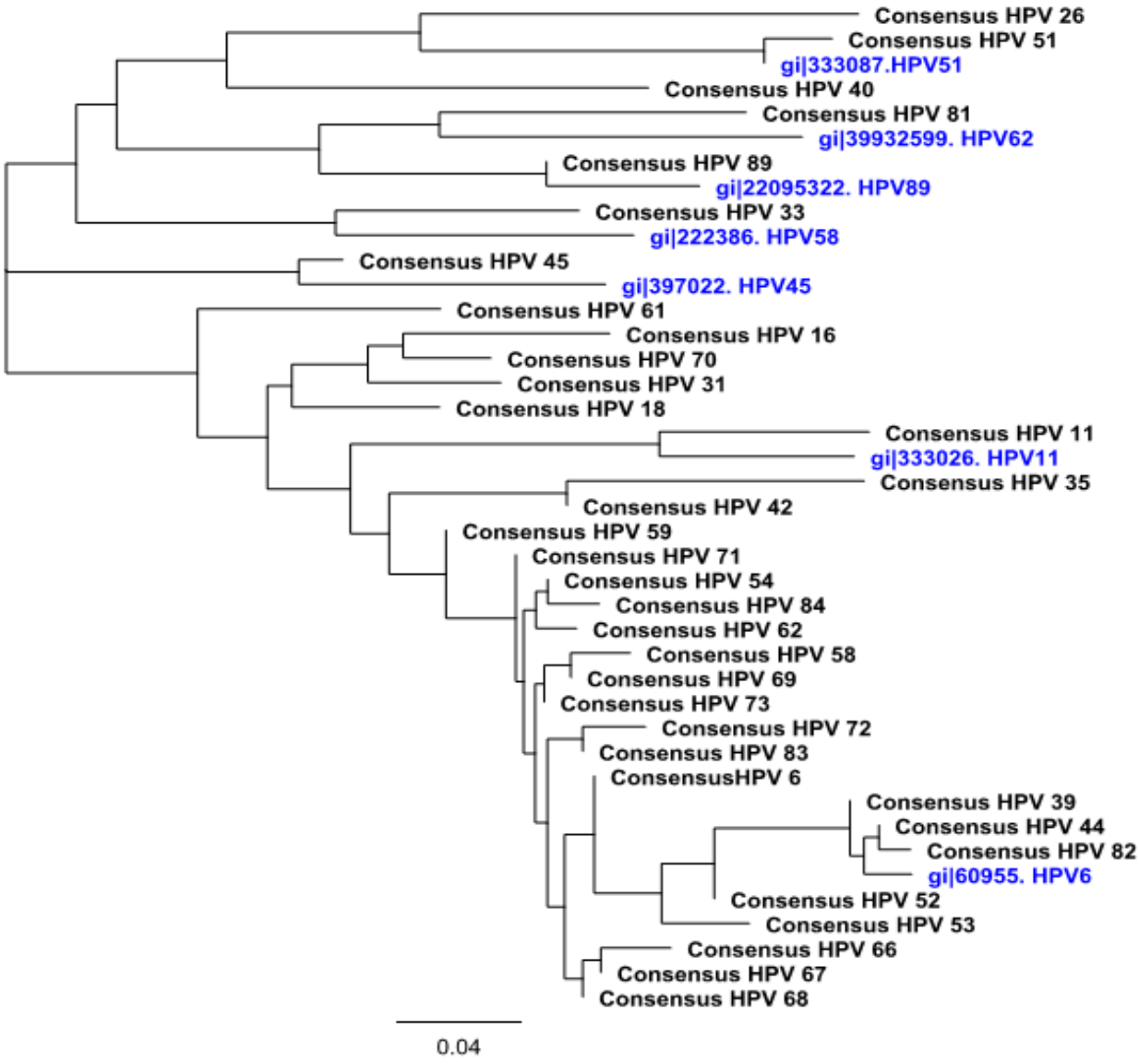

Figure 3.11: A representative phylogenetic tree of sample identity ORHC-016. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-016 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 89 (CP618), 62, 58, 51, 45, 11 and 6 for sample identity ORHC-016, it is important to note that only the detected genotypes were used to compute this tree.

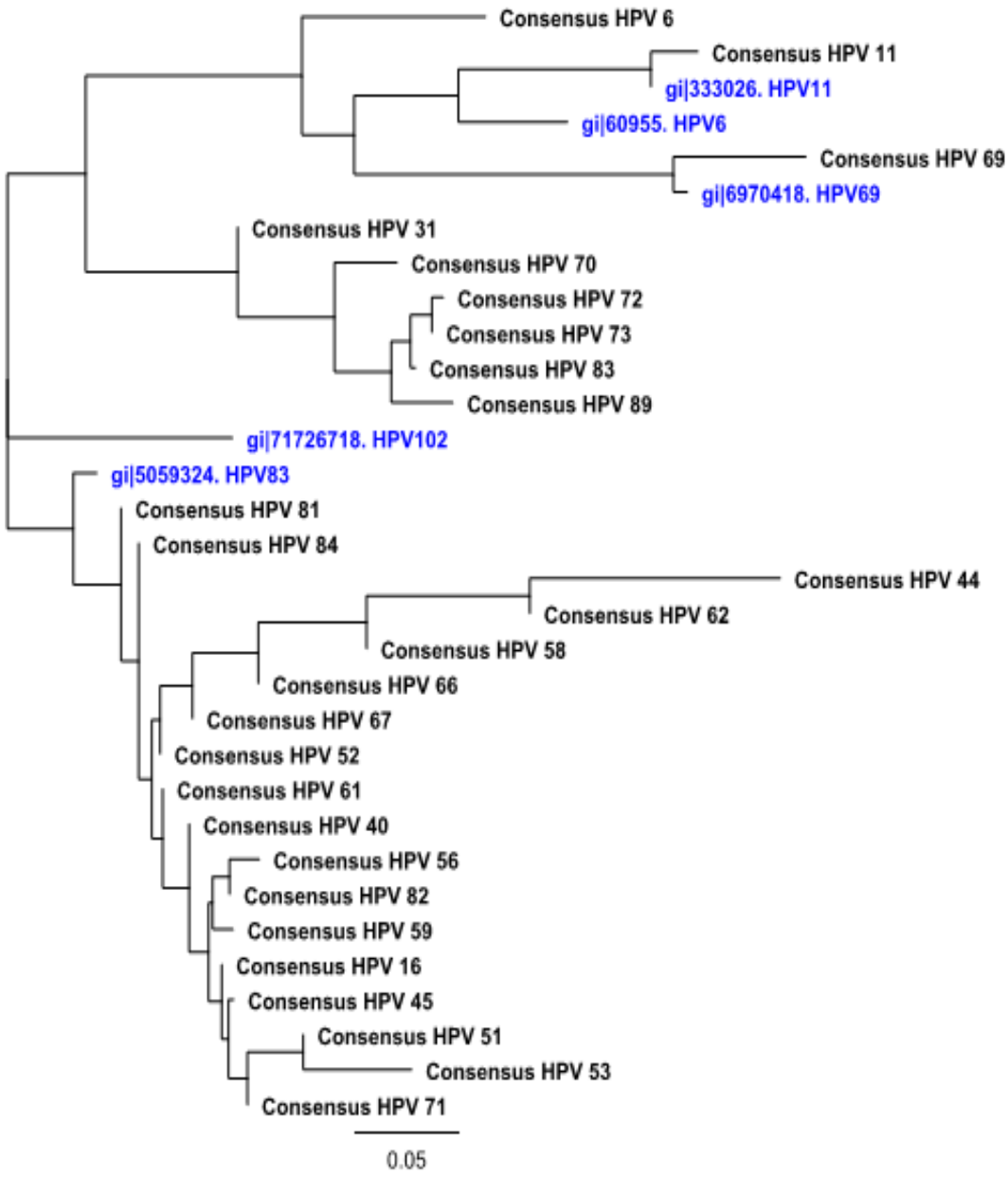

Figure 3.12: A representative phylogenetic tree of sample identity ORHC-028. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-028 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 102, 83, 69, 11 and 6 for sample identity ORHC-028, it is important to note that only the detected genotypes were used to compute this tree.

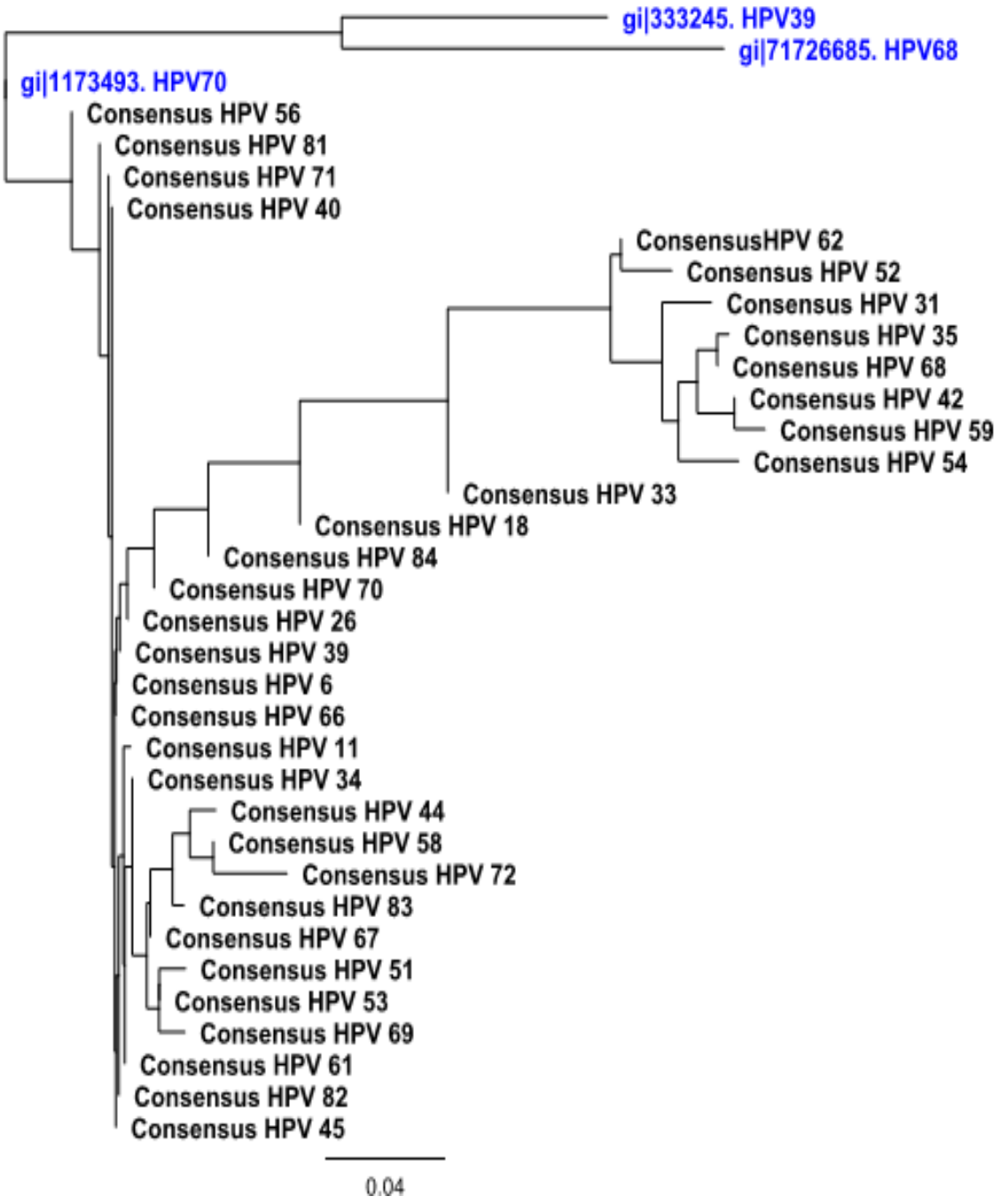

Figure 3.13: A representative phylogenetic tree of sample identity ORHC-031. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-031 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 70, 68 and 39 for sample identity ORHC-031, it is important to note that only the detected genotypes were used to compute this tree.

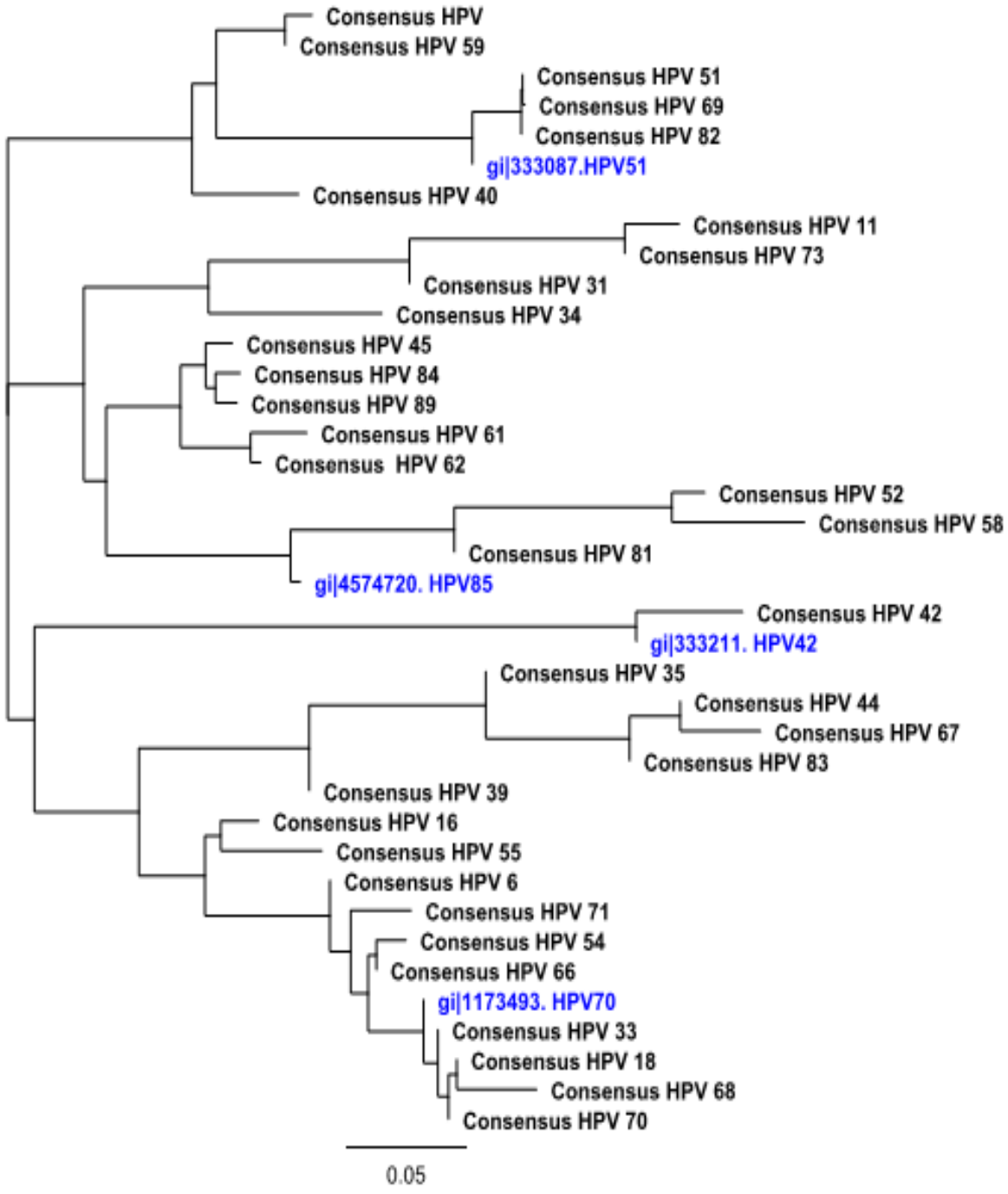

Figure 3.14: A representative phylogenetic tree of sample identity ORHC-040. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-040 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 85, 70, 51 and 42 for sample identity AHDR-040, it is important to note that only the detected genotypes were used to compute this tree.

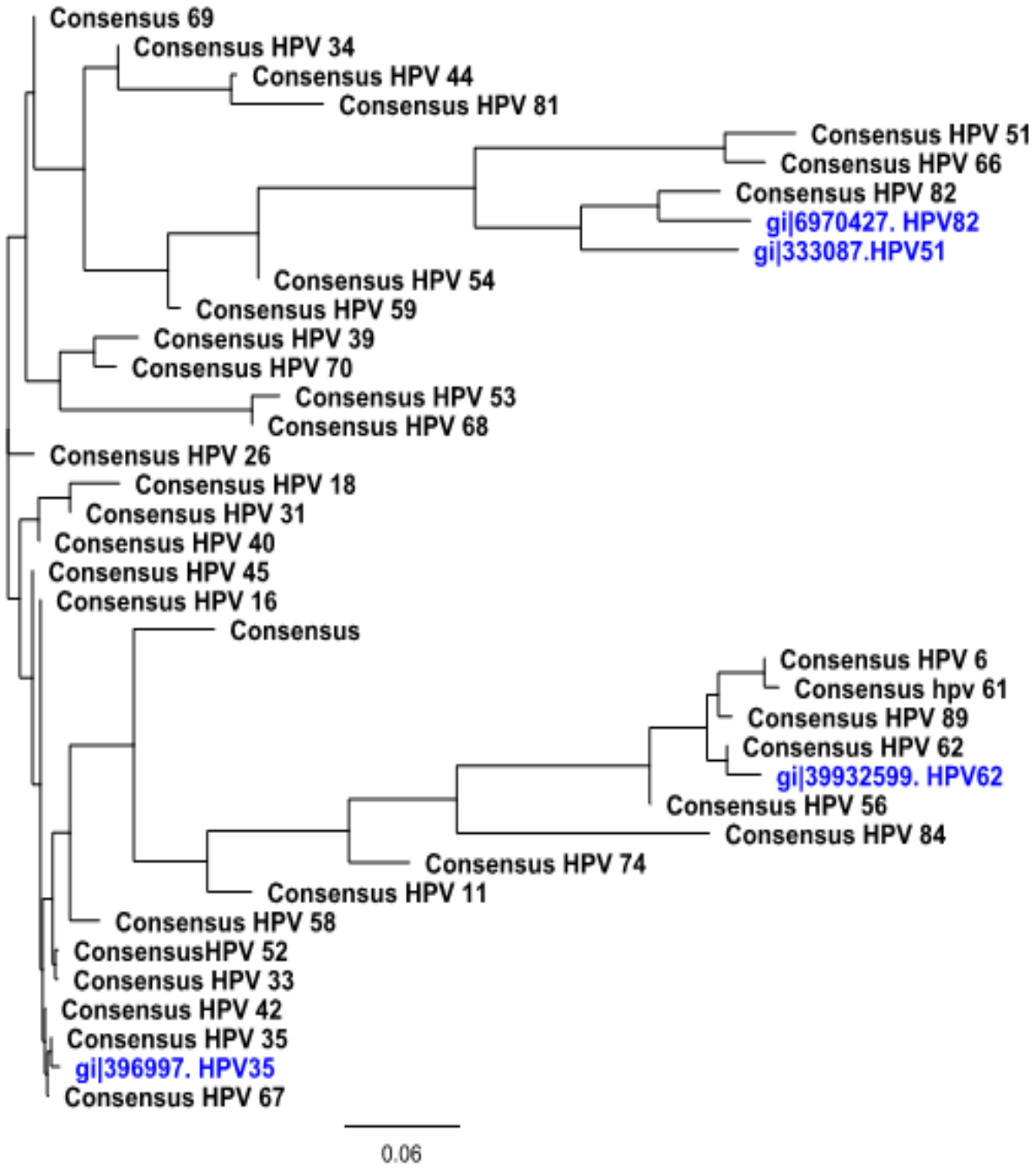

Figure 3.15: A representative phylogenetic tree of sample identity ORHC-043. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-043 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 82, 62, 51 and 35 for sample identityORHC-043, it is important to note that only the detected genotypes were used to compute this tree.

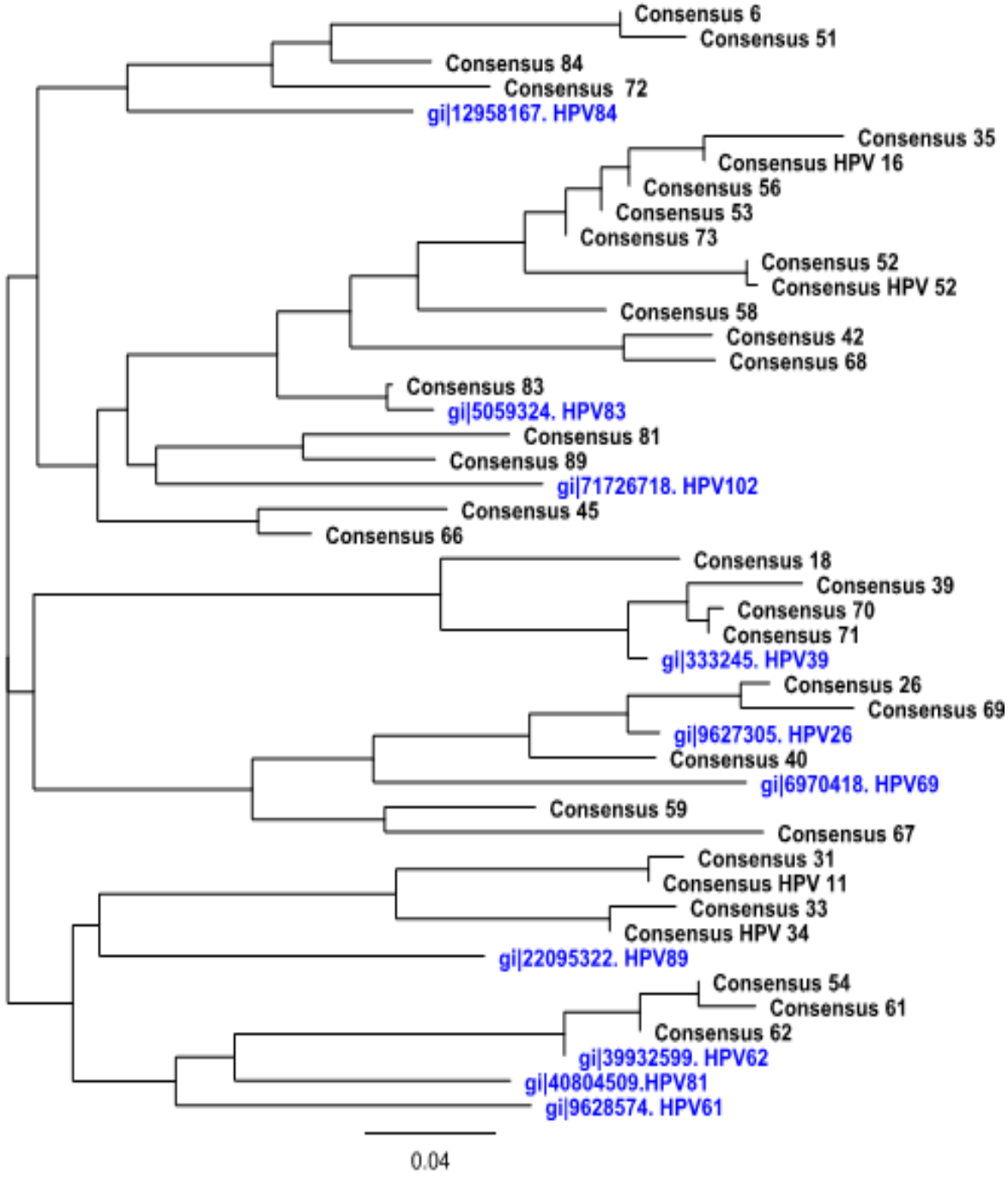

Figure 3.16: A representative phylogenetic tree of sample identity ORHC-046. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping ORHC-046 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 102, 89 (CP6108), 84, 83, 81, 69, 62, 61, 39 and 26 for sample identityORHC-046, it is important to note that only the detected genotypes were used to compute this tree.

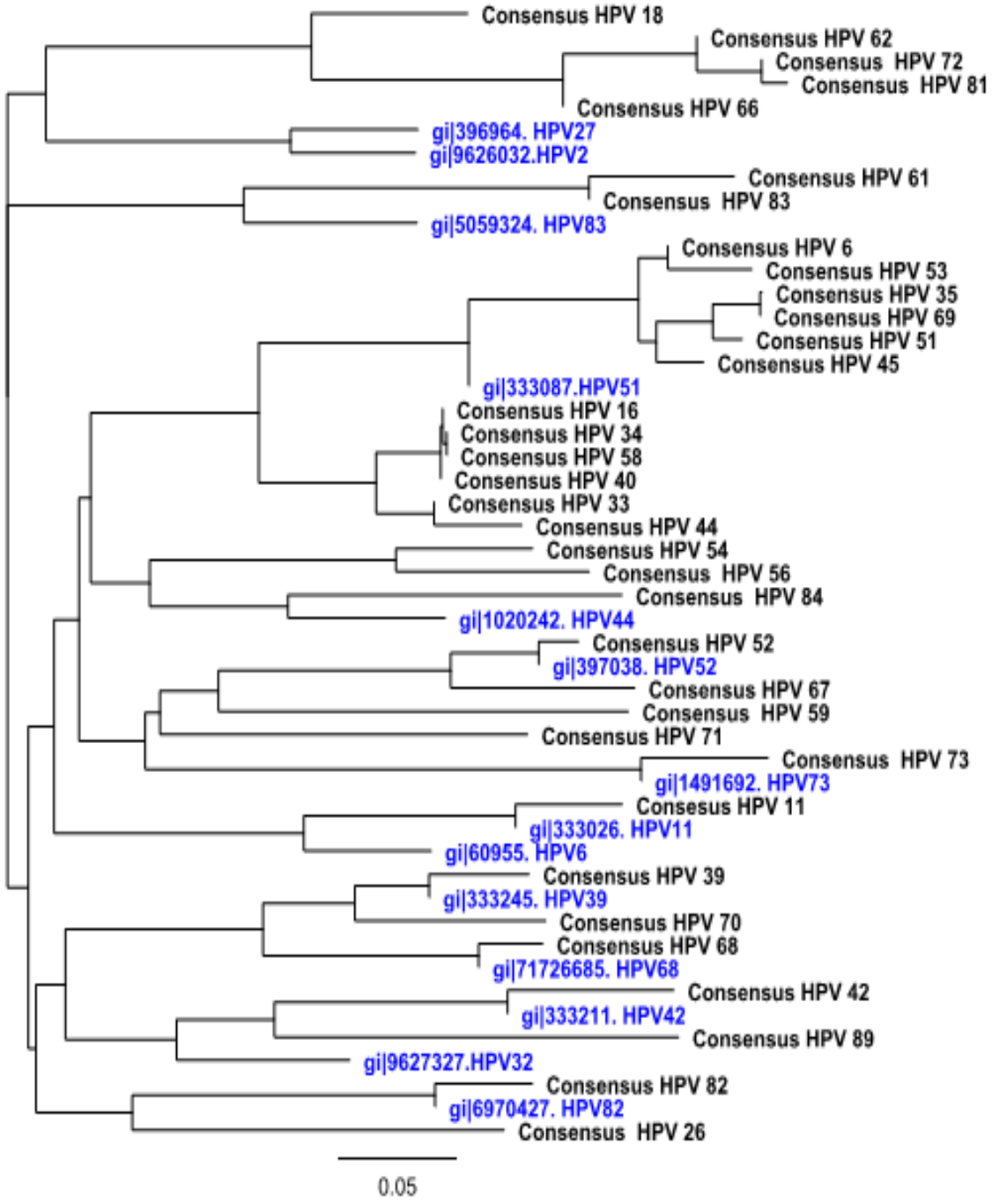

Figure 3.17: A representative phylogenetic tree of sample identity OSHC-007. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping OSHC-007 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 83, 82, 73, 68, 52, 51, 44 (55), 42, 39, 32, 27, 11, 6 and 2 for sample identity OSHC-007, it is important to note that only the detected genotypes were used to compute this tree.

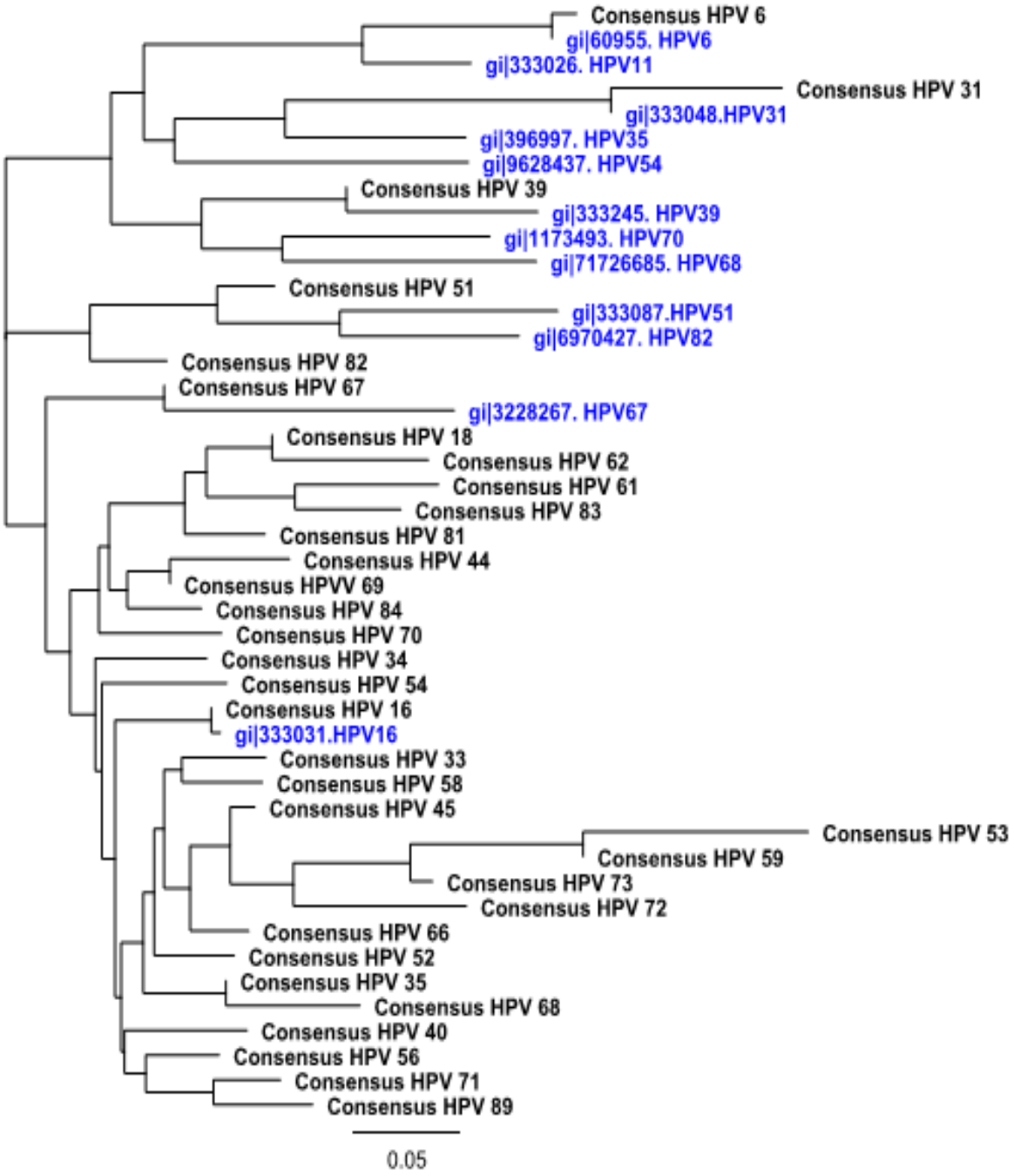

Figure 3.18: A representative phylogenetic tree of sample identity OTHC-014. To compute the tree, geneious tree builder neighbour-joining method was used. The black sequences denotes all the generated consensus sequences extracted after mapping OTHC-014 with various reference sequences belonging to L1 *Alpha* species of HPV downloaded from Papilloma virus Episteme (PaVE) database. It should be noted that the tree was computed using only sequences which contigs were found, sequences which no contigs were found were considered as unmatched and were excluded when computing the tree. The sequences in blue illustrate the detected genotypes namely HPV types 82, 70, 68, 67, 54, 51, 39, 35, 31, 16, 11 and 6 for sample identity OTHC-014, it is important to note that only the detected genotypes were used to compute this tree.

Next Generation sequence assay identified all the HPV types detected by LA assay despite the fact that genotyping data was only available for 15 study subjects. Ten HPV genotypes that were not picked up by LA array on the same samples and that include HPV type: 11, 82, 33, 31, 55, 64, 40, 56, 58 and 67 were picked by NGS. It should be noted that HPV type 55 and 64 on the genotyping reference guide are presently called the subtypes of 44 and 34 individually since they've lost their type status; HPV type CP6108 is also called HPV 89. Genotyping data was successfully generated for all the 20 positive amplicons by PCR. Moreover, 6 additional HPV genotypes namely type 102, 85, 32, 30, 27 and 2 not incorporated in the Roche Linear Array Genotyping test were detected by NGS technology in this study.

3.7.2 DISTRIBUTION OF SINGLE AND MULTIPLE INFECTIONS

Fourteen of the 15 women (93%) in whom HPV DNA was detected by PCR had multiple HPV infections by NGS technology. Only one woman was detected with single infection. Table 3.3 shows the distribution of single and multiple infections detected in the study population.

Table 3.3: Distribution of single and multiple infections by NGS and phylogenetic analysis.

Study subjects	Genotypes detected	HIV status
1. AHDR-S117	❖ HPV types 18, 32, 33, 51, 52, 53, 58, 66, 67, 71 and 82	Positive
2. AHDR-R297	Not successfully sequenced	Positive
3. AHDR-R300	❖ HPV types 18, 64, 59 and 73	Positive
4. AHDR-R308	Not successfully sequenced	Positive
5. ORHC-005	❖ HPV types 11, 16, 51, 53 and 84	Positive
6. ORHC-006	❖ HPV types 30, 33, 53, 58 and 68	Positive
7. ORHC-007	❖ HPV type 71	Positive
8. ORHC-012	❖ HPV types 6, 11, 35, 56, 58, 61, 62, 68, 70, 71 and CP6108	Positive
9. ORHC-014	❖ HPV types 6, 11, 61, 66 and 72	Positive
10. ORHC-016	❖ 6, 11, 45, 51, 58, 62 and CP6108	Positive
11. ORHC-028	❖ HPV types 6, 11, 69, 83 and 102	Positive
12. ORHC-031	❖ HPV types 39, 68 and 70	Positive
13. ORHC-032	Not successfully sequenced	Positive
14. ORHC-040	❖ HPV types 42, 51 70 and 85 `	Positive
15. ORHC-043	❖ HPV types 35, 51, 62 and 82	Positive
16. ORHC-046	❖ HPV types 26, 39, 61, 62, 69, 81, 83, 84, 102 and CP6108	Positive
17. OSHC-007	❖ HPV types 2, 6, 11, 27, 39, 51, 52 , 55, 42, 32, 68, 73, 82 and 83	Negative
18. OSHC-008	❖ Not successfully sequenced	Negative
19. OSHC-010	❖ Not successfully sequenced	Negative
20. OTHC-014	❖ HPV types 6, 11, 16, 31, 35, 39, 51, 54,67, 68, 70, and 82	Negative

Thirteen of the 15 women were HIV positive. The distribution of HPV infection among the study subjects was as follows: single and triple infection was detected in one participant (7%) respectively; most participants (87%) harboured more than 3 HPV types.

3.7.3 DISTRIBUTION OF LR HPV INFECTING GENOTYPES

In contrast to LA assay, from the 15 (100%) successfully sequenced genotyping data, NGS detected 22 LR HPV infecting genotypes namely: types 2, 6, 11, 27, 30, 32, 40, 42, 54, 55, 61, 62, 64, 70, 71, 72, 81, 83, 84, 85, 102 and CP6108. In addition, this includes 4 genotypes not picked up by the LA kit and 6 infecting genotypes not incorporated in the LA. The most frequent detected genotype under LR HPVs in increasing order of frequency included HPV type 62 and 70 (27%), 6 (40%) and 11 (47%). HPV types 40, 42, 54, 72, 64, and 81 were the least detected genotypes with n=1 (7%) each. Furthermore, the common combination observed among the participants was type 6 and 11.

3.7.4 DISTRIBUTION OF HR HPV INFECTING GENOTYPES

A total of 20 HR HPV types were detected in 14 of 15 participants (93%), and that includes HPV types 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67, 68, 69, 73 and 82. It should be noted that HR HPV types 31, 33, 56, 58, 67 and 82 were only detected genotypes by NGS technology. The most frequent detected genotypes in the study population by NGS in increasing order of frequency include type 35 (21%), 39 and 82 (29%) each, 68 (36%) and 51 (50%). In addition, HPV types 26, 31, 45, 53, 56, 58 and 66 were the least detected genotypes n=1 (7%) in the study population. HPV 39 and 68 were observed as the common combination detected under HR HPV types.

3.8 COMPARISON OF DETECTED GENOTYPES BY LA AND NGS

Overall, in this study population, 43 LR and HR HPV types were detected by both assays. However, 10 HPV genotypes namely types 11, 31, 33, 40, 55, 56, 58, 64, 67, and 82 were only detected by NGS technology and all these are the genotypes incorporated in the LA genotyping test kit however, the commercial kit failed to pick them up. Furthermore, NGS identified 6 additional genotypes namely: HPV types 2, 27, 30, 35, 85 and 102 not incorporated in the Roche LA genotyping test kit. Refer to table 3.4 for the detected and undetected genotypes by both assays. More multiple infections were detected by NGS method.

Table 3.4: Detected genotypes by LA and NGS assay

	LA assay	NGS assay
HPV types		
2	*	✓
6	✓	✓
11	*	✓
16	✓	✓
18	✓	✓
26	✓	✓
27	*	✓
30	*	✓
31	*	✓
32	*	✓
33	*	✓
35	✓	✓
39	✓	✓
40	*	✓
42	✓	✓
45	✓	✓
51	✓	✓
52	✓	✓
53	✓	✓
54	✓	✓
55	*	✓
56	*	✓
58	*	✓
59	✓	✓
61	✓	✓
62	✓	✓
64	*	✓
66	✓	✓
67	*	✓
68	✓	✓
69	✓	✓
70	✓	✓
71	✓	✓
72	✓	✓
73	✓	✓
81	✓	✓
82	*	✓

83	✓	✓
84	✓	✓
85	*	✓
CP6108	✓	✓
IS39 (subtype of HPV 82)	✓	✓
102	*	✓

Summary of all the detected genotypes by Roche LA and NGS

KEY: The black checkmarks denote the detected HPV genotypes by both methods, while the star marks represent HPV genotypes detected only by NGS method. It should be noted that the non-bolded genotypes detected by NGS were the ones that were missed by LA method. However, the bolded genotypes denote additional genotypes identified by NGS method that are not incorporated in the Roche LA genotyping test kit.

3.9 OVERVIEW OF THE 20 CASES POSITIVE BY PCR SUBJECTED TO STATISTICAL ANALYSIS

From the 20 HPV positive cases (women), the age range was n=8 (40%). Most women n=16 (80%) had only one sexual partner, n=13 (65%) were single with n=11 (65%) of the women falling between the age range of 31-40 of age at first sexual intercourse. With regards to the highest level of education, n=13 (65%) had grade 12 and less then followed by those who reported to have had a diploma n=6 (30%) with 60% (n=12) of employment rate. Furthermore, almost all n=19 (95%) of the women reported not ever been vaccinated against HPV infection, and n=8 (40%) reported to have had Pap smear testing once or twice in their life time respectively. In addition, n=16 (80%) and 4 (20%) of the women were HIV positive and negative respectively.

3.10 RELATIONSHIP BETWEEN HPV POSITIVE STATUS AND ASSOCIATED RISK FACTORS

After genotyping, the demographic and clinical data of all the 20 cases positive by PCR were subjected to statistical analysis (Chi-square) in order to determine if there was a relationship between HPV positive status and the associated risk factors. Chi-square test wasn't performed for gender since this study comprised only women so it was constant variable throughout. Among the 20 HPV positive cases the following risk factors were found to be statically insignificant: age, occupation, pap smear testing status in a life time and HIV status with a p-value of p=0.423, 0.371, 0.055

and 0,077 respectively this suggests no positive relationship was found between HPV positive status and the above mentioned factors. In contrast; the findings on smoking status (p=0.000), age at first sexual intercourse (p=0.011), vaccination status (p=0.000), gender of sexual partner (p=0.000), highest level of education (p=0.004), marital status (p=0.008) and number of sexual partners (p=0.000) showed a positive relationship with statistical significance.

CHAPTER 4: DISCUSSION, LIMITATIONS AND CONCLUSION OF THE STUDY

Despite the availability of vaccination campaigns that the South African government has implemented against initial HPV infection, cervical cancer still remains one of the public health problems among women in South Africa. In addition, lack of knowledge about HPV infection and cervical cancer screening among young and old adults still remains a challenging and controversial issue especially in less developed countries such as South Africa. This might be due to limited resources, lack of proper health care practices and lack of health education regarding HPV infection and its disease progression. Among the 20 HPV positive cases, the age range was 24-60, 19 (95%) women who participated in this current study reported never been vaccinated against HPV infection, with n=16 (80%) reported to have had Pap smear testing once/twice in their life time. Thus, HPV the causative agent of cervical cancer (CC) still remains one of the leading major public health issues.

The current study aimed to determine the burden and genetic diversity of HPV infecting genotypes in a cohort of women in routine care in northern South Africa.

The Linear array (LA) polymerase chain reaction (PCR) was employed to amplify the L1 HPV gene comprising of 450bp. Cervical specimens were collected from a cohort of 87 women living with/out HIV who were visiting the Health care centers as part of their routine care in the northern South Africa.

Studies show that the distribution of HPV infection is influenced by distinct factors such as geographical location (Anderson et al., 2013; Quek et al., 2013) and also the risk factors for acquiring the infection. Therefore, it is imperative to develop future vaccines that are specific for a particular geographical location. And to do so, it is critical to first screen and determine the HPV genotypes prevalent in that particular location.

The results of this study document the prevalence of 23% (n=20) of HPV DNA infection, of the twenty, 16 (80%) were living with HIV. Despite the apparent difference in the distribution of HPV infection between HIV positive and negative individuals, these findings illustrate that HPV infection can be found in both populations.

The findings of this study are in discordant with the high HPV prevalence of 76.3% reported in Kwazulu-Natal by Ebrahim et al., (2016) which is almost 3 times higher compared to the reported prevalence by this study. One of the many reasons for this discordance might be because in the present study HPV infection was screened not only from the HIV positive individuals where the prevalence of HPV is very high. Another study that was conducted in Gauteng Province also reported a high prevalence of 74.6% which had almost similar prevalence to the study reported by Ebrahim et al., (2016) (Richter et al., 2013). However, the prevalence reported in this study was in concordant with the study carried out in Cape-Town Province by McDonald et al., (2014), the study reported the overall prevalence of 25% which is more like similar to the 23% obtained in this present study. Another study conducted in United States of America by Dunne et al., (2007) also reported the overall prevalence of 27% which is in harmony with the present study.

Furthermore, the positive PCR amplicons were assayed first for 37 HPV types incorporated on the Roche LA reverse line blot assay for genotyping. The findings illustrated the overall proportion of 73% (27 of 37) of LR and HR HPV genotypes and that include types 84, 83, 81, 73, 72, 71, 70, 69, 68, 66, 62, 61, 59, 54, 53, 52, 51, 45, 42, 39, 35, 26, 18, 16, 6, IS39 and CP6108 observed to be detected in at least one of the study participants, showing the diversity of HPV types present in a cohort of women in routine care. However, Roche LA method didn't detect 10 HPV types incorporated in the test namely types 82, 67, 64, 58, 56, 55, 40, 33, 31, and 11 in any of the study isolates.

With regards to HPV infection, the distribution of multiple HPV genotypic infection by LA was found to be high and it was observed to be harboured by 75% (15 of 20) of the women who participated in the current study. These findings are in agreement with the results reported by Mbulawa et al., (2015), where the study documented a high prevalence of multiple genotypic HPV infection compared to single infection in their study population. A study conducted by Trottier et al., (2008), reported that infection with HPV multiple types is associated with a greater rate of persistent HPV infection compared to single infection. In addition, Munagala et al., (2009) illustrated that individuals with multiple HPV infections compared to the ones harbouring single infections, they have an increased risk of having a larger growth and also poorer

response when it comes to the treatment of cancer. In addition, a study conducted by Thomas et al., (2000) reported that one of the crucial risk factors for harbouring multiple HPV genotypic infection is sexual behaviour like frequency of intercourse or the characteristics of the sexual partner (numbers of HPV types in the partners).

Still using LA, the proportion of women harbouring LR HPV types considered as non-oncogenic was detected in 14 (70%) women and that include 13 LR types namely: types 84, 83, 81, 72, 71, 70, 62, 61, 54, 42, 6, IS39 and CP6108. The most common detected genotype in increasing order of frequency under LR types was HPV types 81 (43%), 62 (36%), 83 and 61 (21%). The findings reported in this current study was similar to the one reported by Zhou et al., in which HPV 81 was documented to be the most prevalent genotype (Zhou et al., 2017). Furthermore, this study documented HPV types 71, 54, 6 and IS39 as the least (7%) LR detected genotypes. However, the most common LR types 6 and 11 responsible for causing genital warts were not common at all by LA. In the current study LR HPV type 11 documented by many studies as a common HPV genotype under LR types wasn't detected at all by the Roche LA method, in addition HPV 6 together with other genotypes were among the least detected types.

In contrast to LR HPV types, 14 oncogenic types grouped as HR HPV types were detected in 17 women with the proportion of 85% and included types 73, 69, 68, 66, 59, 53, 52, 51, 45, 39, 35, 26, 18 and 16, with at least one woman harbouring one or more of the detected genotypes. This proportion of women detected with HR genotypes was high compared to the one of 79% reported by McDonald et al., 2014. The most common genotypes observed in order of decreasing frequency under HR was HPV type: 66, 53, 52, 51, 45, 18 and 16 (18%) respectively and HPV type: 73, 68, 59, 39 and 35 harboured by 12% of the participants respectively. The least detected genotype under HR with 6% was HPV type 26 and 69. According to a study reported by Zhang et al., they documented HPV 52, 16, 58, 18 and 33 as the five most common HR HPV types in descending order (Zhang et al., 2013). Another study by Pitta et al., documented HPV types 52, 51, 18 and 16 as some of the most prevalent HR genotypes detected in their study population (Pitta et al., 2009) which is more or less similar to the results documented by this study.

HR HPV genotypes especially the persistence infection of type 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59 and 66 have been documented to be highly carcinogenic, with more than 70% of cervical cancer cases are known to be related to HPV 16 and 18 (Munoz et al., 2003; de Villers et al., 2004; IARC, 2012). In this current study although HPV 16 and 18 was the most common types with a frequency of 18% each, no study participants were seen to be harbouring the combination of the two most common HPV types under HR types covered by Bivalent vaccine. The detected HR HPV genotypes were types 73, 69, 68, 66, 59, 53, 52, 51, 45, 39, 35, 26, 18 and 16. Although not all the genotypes incorporated on the kit were identified by the Roche LA, this method showed that there is a diversity of both LR and HR HPV genotypes prevalent in the cohort of women in routine care.

To validate the HPV findings detected by Roche LA method and also to identify extra additional genotypes not detected by LA, NGS was performed using the other portion (undenatured amplicons) of the same 20 PCR positive amplicons. It is critical to mention that all the reference sequences belonging to L1 Alpha species were used for mapping. The reference sequences were downloaded from the Papillomavirus Episteme (PaVE). From the 20 successfully amplified PCR amplicons, genotyping data using NGS was successfully generated from 15 isolates (75%).

From the 15 genotyped findings, the proportion of women with multiple HPV infections was found to be 93% (14 of 15) which is high compared to the proportion this study documented by LA. Overall, NGS further detected 10 more HPV types and that includes types 11, 82, 33, 31, 55, 64, 40, 56, 58 and 67. Furthermore, 6 additional genotypes were identified by NGS and that includes type 102, 85, 32, 30, 27 and 2, it is critical to mention that this 6 genotypes are not incorporated in the Roche LA test. The additional identified genotypes especially HPV types 102 and 32 are in agreement with the study conducted by Flores-Miramontes et al., (2015) where they identified 5 additional genotypes by NGS technology not incorporated in the Roche LA test. Another study by Nowak et al., reported an additional HPV genotypes by NGS and that included HPV 30 and 32 and other 8 genotypes, these genotypes were not among the 37 HPV types detected by the Roche LA method,

making HPV 30 and 32 to be in agreement with the additional genotypes detected in this study (Nowak et al., 2017).

Contrary to Roche LA method, NGS technology identified 22 LR HPV types. Among which it showed more sensitivity to HPV types 6 (40%) and 11 (47%) which is in disagreement to the LA findings documented by this study. The reason of this might be HPV 6 and 11 among the study participants were present as the minority populations hence the commercial kit fails to pick them up in high quantity. Using NGS the most common LR type 81 by LA was detected as the least among with other 5 genotypes with the proportion of 7%.

Under the HR group, the detected genotypes included HPV types namely: types 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 67, 68, 69, 72 and 82. These results were in harmony with the ones reported by Meisal et al., (2017). In their study using NGS, the detected HR HPV genotypes were types 16, 18, 30, 33, 35, 39, 40, 42, 43, 45, 51, 52, 53, 54, 56, 58, 62, 66, 69, 73, 89, and 90, the proportion of women harbouring HR HPV types was 93%. It should be noted that in the current study an extra 6 HR HPV types 31, 33, 56, 58, 67 and 82 were only detected by NGS technology, implying that Roche LA method wasn't sensitive enough to identify the mentioned HPV genotypes although they are incorporated in the Roche LA, a reason for this might be these genotypes formed part of the minority population hence were only identified by NGS. Furthermore, if NGS wasn't employed those additional genotypes could have been missed. Therefore, it is of importance to use both techniques. The most prevalent genotypes under HR types was 35 (21%), 39; 56 and 82 (29%), 68 (36%) and 51 (50%) in increasing order of frequency. The least detected genotypes were HPV 45, 66, 56, 26, 31, 53 and 58.

Overall, NGS technology was more sensitive in detecting a wide range of diversity of HPV genotypes prevalent in a cohort of women in routine care compared to Roche LA genotyping test. Therefore, more studies on screening and genotyping HPV infection in northern South Africa using various screening and genotyping tools are needed so that they can help to clarify the viral diversity in the population. More importantly, it should be noted that this present study detected all HR HPV

genotypes known to be highly carcinogenic which are targeted by the Gardasil-9 vaccine not yet in use in South Africa. The nonavalent targets HPV types 6, 11, 16, 18, 31, 33, 45, 52 and 58, with the first two LR associated with causing genital warts and the last seven genotypes are known to cause cervical cancer if they become persistent. Which means if this vaccine is not introduced in the near future we will end up having a case where we have a lot of women having persistent HR infections that can thereafter in the future leads to the development cervical cancer. In addition since the study documented the presence of HR types there is a chance that these already infected women can transmit the detected genotypes to their sexual partners, as a result we might end up having a population where there is an uncontrolled circulation of cancerous genotypes. It is therefore strongly encouraged that the South African government consider the implementation of this vaccine in the future in order to control HPV infection in northern South Africa.

Following genotyping by both methods, the demographic and clinical data of all the 20 cases positive by PCR were subjected to statistical analysis in order to determine if there was a relationship between HPV positive status and the associated risk factors. The statistical findings were in harmony with the results reported by Bahmanya et al., (2012), the study found a statistically association on the behavioural risk factors such as smoking, marital status, age at first sexual intercourse and number of sexual partners.

LIMITATIONS OF THE STUDY: This study comprised a relative small sample size with only positive PCR amplicons by LA subjected to hybridization, genotyping by LA and NGS assays. As a result some of the genotypes might have been present on the negative PCR amplicons. Furthermore, this study targeted only the black African women, while the mixed population might have highlighted a wide range of variations compared to a single race.

Therefore, it is strongly recommended that a broader study be conducted using various commercial kits and other techniques to all races in order to compare the burden and diversity of more genotypes of HPV not covered by the commercial kits. However, this study documents the presence of HPV infection, its diversity, the prevalent genotypes detected by the LA assay and NGS in the study cohort and also the association between HPV positive status and associated risk factors.

CONCLUSION: Amplification of HPV DNA from cervical specimens demonstrated the presence of HPV infection in the study cohort, with an infection rate of 23%. The findings suggest that there is a diversity of HPV genotypes prevalent in the study population by Roche LA and NGS methods. Next generation sequencing approach was observed to be more sensitive in detecting the HPV genotypes. This current study also documented a positive association between HPV positive DNA and smoking, age at first sexual intercourse, marital status, vaccination status, gender of sexual partner, highest level of education and number of sexual partners.

REFERENCES

Adam, E., Berkova, Z., Daxnerova, Z., Icenogle, J., Reeves, W.C. and Kaufman, R.H., 2000. Papillomavirus detection: demographic and behavioral characteristics influencing the identification of cervical disease. *American Journal of Obstetrics and Gynecology*, *182*(2), pp.257-264.

Adler, D.H., Wallace, M., Bennie, T., Mrubata, M., Abar, B., Meiring, T.L., Williamson, A.L. and Bekker, L.G., 2014. Cervical dysplasia and high-risk human papillomavirus infections among HIV-infected and HIV-uninfected adolescent females in South Africa. *Infectious Diseases in Obstetrics and Gynecology*, *2014*.

Anderson, L., O'rorke, M., Jamison, J., Wilson, R., Gavin, A. and HPV Working Group members, 2013. Prevalence of human papillomavirus in women attending cervical screening in the UK and Ireland: New data from northern Ireland and a systematic review and meta-analysis. *Journal of Medical Virology*, *85*(2), pp.295-308.

Ansorge, W.J., 2009. Next-generation DNA sequencing techniques. *New Biotechnology*, *25*(4), pp.195-203.

Baay, M.F., Q`uint, W.G., Koudstaal, J., Hollema, H., Duk, J.M., Burger, M.P., Stolz, E. and Herbrink, P., 1996. Comprehensive study of several general and type-specific primer pairs for detection of human papillomavirus DNA by PCR in paraffin-embedded cervical carcinomas. *Journal of Clinical Microbiology*, *34*(3), pp.745-747.

Bahmanyar, E.R., Paavonen, J., Naud, P., Salmerón, J., Chow, S.N., Apter, D., Kitchener, H., Castellsagué, X., Teixeira, J.C., Skinner, S.R. and Jaisamrarn, U.,

2012. Prevalence and risk factors for cervical HPV infection and abnormalities in young adult women at enrolment in the multinational PATRICIA trial. *Gynecologic Oncology*, *127*(3), pp.440-450.

Baker, T.S., Newcomb, W.W., Olson, N.H., Cowsert, L.M., Olson, C. and Brown, J.C., 1991. Structures of bovine and human papillomaviruses. Analysis by cryoelectron microscopy and three-dimensional image reconstruction. *Biophysical Journal*, *60*(6), pp.1445-1456.

Bansal, A., Singh, M.P. and Rai, B., 2016. Human papillomavirus-associated cancers: A growing global problem. *International Journal of Applied and Basic Medical Research*, *6*(2), p.84.

Bartlett, J.M. and Stirling, D., 2003. A short history of the polymerase chain reaction. In *PCR protocols* (pp. 3-6). Humana Press.

Bauer, H.M., Ting, Y., Greer, C.E., Chambers, J.C., Tashiro, C.J., Chimera, J., Reingold, A. and Manos, M.M., 1991. Genital human papillomavirus infection in female university students as determined by a PCR-based method. *Journal of American Medical Association*, *265*(4), pp.472-477.

Bernard, H.U., 2005. The clinical importance of the nomenclature, evolution and taxonomy of human papillomaviruses. *Journal of Clinical Virology*, *32*, pp.1-6.

Burk, R.D., Kelly, P., Feldman, J., Bromberg, J., Vermund, S.H., Dehovitz, J.A. and Landesman, S.H., 1996. Declining prevalence of cervicovaginal human papillomavirus infection with age is independent of other risk factors. *Sexually Transmitted Diseases*, *23*(4), pp.333-341.

Chaturvedi, A.K., Engels, E.A., Pfeiffer, R.M., Hernandez, B.Y., Xiao, W., Kim, E., Jiang, B., Goodman, M.T., Sibug-Saber, M., Cozen, W. and Liu, L., 2011. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *Journal of Clinical Oncology*, *29*(32), pp.4294-4301.

Clifford, G.M., Rana, R.K., Franceschi, S., Smith, J.S., Gough, G. and Pimenta, J.M., 2005. Human papillomavirus genotype distribution in low-grade cervical lesions: comparison by geographic region and with cervical cancer. *Cancer Epidemiology and Prevention Biomarkers*, *14*(5), pp.1157-1164.

Danos, O., Katinka, M. and Yaniv, M., 1982. Human papillomavirus 1a complete DNA sequence: a novel type of genome organization among papovaviridae. *The EMBO Journal*, *1*(2), p.231.

Dempsey, A.F. and Mendez, D., 2010. Examining future adolescent human papillomavirus vaccine uptake, with and without a school mandate. *Journal of Adolescent Health*, *47*(3), pp.242-248.

De Sanjosé, S., Diaz, M., Castellsagué, X., Clifford, G., Bruni, L., Muñoz, N. and Bosch, F.X., 2007. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. *The Lancet infectious diseases*, *7*(7), pp.453-459.

De Villiers, E.M., Fauquet, C., Broker, T.R., Bernard, H.U. and zur Hausen, H., 2004. Classification of papillomaviruses. *Virology*, *324*(1), pp.17-27.

Dickson, E.L., Vogel, R.I., Bliss, R.L. and Downs Jr, L.S., 2013. Multiple-type HPV infections: a cross-sectional analysis of the prevalence of specific types in 309,000 women referred for HPV testing at the time of cervical cytology. *International journal*

of gynecological cancer: official Journal of the International Gynecological Cancer Society, *23*(7).

Dillner, J., 2015, April. Prevention of human papillomavirus–associated cancers. In *Seminars in Oncology* (Vol. 42, No. 2, pp. 272-283). WB Saunders.

Doorbar, J., 2005. The papillomavirus life cycle. *Journal of Clinical Virology*, *32*, pp.7-15.

Dochez, C., Bogers, J.J., Verhelst, R. and Rees, H., 2014. HPV vaccines to prevent cervical cancer and genital warts: an update. *Vaccine*, *32*(14), pp.1595-1601.

Drury, S.E.N., Gough, R.E., McArthur, S. and Jessop, M., 1998. Detection of herpesvirus-like and papillomavirus-like particles associated with diseases of tortoises. *Veterinary Record (United Kingdom)*.

Dunne, E.F. and Markowitz, L.E., 2006. Genital human papillomavirus infection. *Clinical Infectious Diseases*, *43*(5), pp.624-629.

Dunne, E.F., Unger, E.R., Sternberg, M., McQuillan, G., Swan, D.C., Patel, S.S. and Markowitz, L.E., 2007. Prevalence of HPV infection among females in the United States. *Journal of the American Medical Association*, *297*(8), pp.813-819.

Ebrahim, S., Mndende, X.K., Kharsany, A.B., Mbulawa, Z.Z., Naranbhai, V., Frohlich, J., Werner, L., Samsunder, N., Karim, Q.A. and Williamson, A.L., 2016. High burden of human papillomavirus (HPV) infection among young women in KwaZulu-Natal, South Africa. *PloS one*, *11*(1), p.e0146603.

Faridi, R., Zahra, A., Khan, K. and Idrees, M., 2011. Oncogenic potential of Human Papillomavirus (HPV) and its relation with cervical cancer. *Virology Journal*, *8*(1), p.269.

Flores-Miramontes, M.G., Torres-Reyes, L.A., Alvarado-Ruíz, L., Romero-Martínez, S.A., Ramírez-Rodríguez, V., Balderas-Peña, L.M.A., Vallejo-Ruíz, V., Piña-Sánchez, P., Cortés-Gutiérrez, E.I., Jave-Suárez, L.F. and Aguilar-Lemarroy, A., 2015. Human papillomavirus genotyping by Linear Array and Next-Generation Sequencing in cervical samples from Western Mexico. *Virology Journal*, *12*(1), p.161.

Giannoudis, A. and Simon Herrington, C., 2001. Human papillomavirus variants and squamous neoplasia of the cervix. *Journal of Pathology*, *193*(3), pp.295-302.

Gillison, M.L., Chaturvedi, A.K. and Lowy, D.R., 2008. HPV prophylactic vaccines and the potential prevention of noncervical cancers in both men and women. *Cancer*, *113*(S10), pp.3036-3046.

Giuliano, A.R., Papenfuss, M., Schneider, A., Nour, M. and Hatch, K., 1999. Risk factors for high-risk type human papillomavirus infection among Mexican-American women. *Cancer Epidemiology and Prevention Biomarkers*, *8*(7), pp.615-620.

Giuliano, A.R., Botha, M.H., Zeier, M., Abrahamsen, M.E., Glashoff, R.H., Van der Laan, L.E., Papenfuss, M., Engelbrecht, S., Van der Loeff, M.F.S., Sudenga, S.L. and Torres, B.N., 2015. High HIV, HPV, and STI prevalence among young Western Cape, South African women: EVRI HIV prevention preparedness trial. *Journal of Acquired Immune Deficiency Syndromes (1999)*, *68*(2), p.227.

Gomez, D.T. and Santos, J.L., 2007. Human papillomavirus infection and cervical cancer: pathogenesis and epidemiology. *Communicating current research and educational topics and trends in applied microbiology*, *1*, pp.680-8.

Hamid, N.A., Brown, C. and Gaston, K., 2009. The regulation of cell proliferation by the papillomavirus early proteins. *Cellular and Molecular Life Sciences*, *66*(10), pp.1700-1717.

Hausen, H., 1982. Human genital cancer: synergism between two virus infections or synergism between a virus infection and initiating events?. *Lancet*, *320*(8312), pp.1370-1372.

Hazard, K., 2007. Cutaneous Human Papillomaviruses [Thesis]. *Malmo: Lund University*.

Herbst, L.H., Lenz, J., Van Doorslaer, K., Chen, Z., Stacy, B.A., Wellehan Jr, J.F., Manire, C.A. and Burk, R.D., 2009. Genomic characterization of two novel reptilian papillomaviruses, Chelonia mydas papillomavirus 1 and Caretta caretta papillomavirus 1. *Virology*, *383*(1), pp.131-135.

Ho, G.Y., Bierman, R., Beardsley, L., Chang, C.J. and Burk, R.D., 1998. Natural history of cervicovaginal papillomavirus infection in young women. *New England Journal of Medicine*, *338*(7), pp.423-428.

Ho, G.Y., Burk, R.D., Klein, S., Kadish, A.S., Chang, C.J., Palan, P., Basu, J., Tachezy, R., Lewis, R. and Romney, S., 1995. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. *Journal of the National Cancer Institute*, *87*(18), pp.1365-1371.

Hontelez, J.A., Lurie, M.N., Newell, M.L., Bakker, R., Tanser, F., Bärnighausen, T., Baltussen, R. and de Vlas, S.J., 2011. Ageing with HIV in south africa. *AIDS (London, England)*, *25*(13).

https://en.wikipedia.org/w/index.php?title=List_of_municipalities_in_Limpopo&oldid=860431908

Lange, C.E., Favrot, C., Ackermann, M., Gull, J., Vetsch, E. and Tobler, K., 2011. Novel snake papillomavirus does not cluster with other non-mammalian papillomaviruses. *Virology journal*, *8*(1), p.436.

Iftner, T. and Villa, L.L., 2003. Chapter 12: Human papillomavirus technologies. *JNCI Monographs*, *2003*(31), pp.80-88.

Jemal, A., Bray, F., Center, M.M., Ferlay, J., Ward, E. and Forman, D., 2011. Global cancer statistics. *CA: A Cancer Journal for Clinicians*, *61*(2), pp.69-90.

Joshi, M. and Deshpande, J.D., 2010. Polymerase chain reaction: methods, principles and application. *International Journal of Biomedical Research*, *2*(1), pp.81-97.

Justice, A.C., 2010. HIV and aging: time for a new paradigm. *Current HIV/AIDS Reports*, *7*(2), pp.69-76.

Khan, S., Jaffer, N.N., Khan, M.N., Rai, M.A., Shafiq, M., Ali, A., Pervez, S., Khan, N., Aziz, A. and Ali, S.H., 2007. Human papillomavirus subtype 16 is common in Pakistani women with cervical carcinoma. *International Journal of Infectious Diseases*, *11*(4), pp.313-317.

Kiviat, N.B. and Koutsky, L.A. (1993). Specific human papillomavirus types as the causal agents of most cervical intraepithelial neoplasia: implications for current views and treatment. *Journal of the National Cancer Institute*. 85(12),pp.934–935.

Koshiol, J., Lindsay, L., Pimenta, J.M., Poole, C., Jenkins, D. and Smith, J.S., 2009. Persistent Human Papillomavirus Infection and Cervical Neoplasia: A Systematic Review and Meta-Analysis. *Journal of Lower Genital Tract Disease*, *13*(1), p.59.

Koutsky, L.A., Ault, K.A., Wheeler, C.M., Brown, D.R., Barr, E., Alvarez, F.B., Chiacchierini, L.M. and Jansen, K.U., 2002. A controlled trial of a human papillomavirus type 16 vaccine. *New England Journal of Medicine*, *347*(21), pp.1645-1651.

Lacey, C.J. (2005). Therapy for genital human papillomavirus related disease. *Journal of Clinical Virology. 32(1),pp.*S83-S90.

Lazarczyk, M., Cassonnet, P., Pons, C., Jacob, Y. and Favre, M., 2009. The EVER proteins as a natural barrier against papillomaviruses: a new insight into the pathogenesis of human papillomavirus infections. *Microbiology and Biology Reviews.* 73(2),pp.348-370.

Lomalisa, P., Smith, T. and Guidozzi, F., 2000. Human immunodeficiency virus infection and invasive cervical cancer in South Africa. *Gynecologic Oncology*, *77*(3), pp.460-463.

Maufort, J.P., Shai, A., Pitot, H.C. and Lambert, P.F., 2010. A role for HPV16 E5 in cervical carcinogenesis. *Cancer Research*, *70*(7), pp.2924-2931.

Mbulawa, Z.Z., Coetzee, D. and Williamson, A.L., 2015. Human papillomavirus prevalence in South African women and men according to age and human immunodeficiency virus status. *BMC Infectious Diseases*, *15*(1), p.459.

McDonald, A.C., Tergas, A.I., Kuhn, L., Denny, L. and Wright, T.C., 2014. Distribution of human papillomavirus genotypes among HIV-positive and HIV-negative women in Cape Town, South Africa. *Frontiers in Oncology*, *4*, p.48.

Meisal, R., Rounge, T.B., Christiansen, I.K., Eieland, A.K., Worren, M.M., Molden, T.F., Kommedal, Ø., Hovig, E., Leegaard, T.M. and Ambur, O.H., 2017. HPV Genotyping of Modified General Primer-Amplicons Is More Analytically Sensitive and Specific by Sequencing than by Hybridization. *PloS one*, *12*(1), p.e0169074.

Molecular biologists for Oncologists 2nd edition. Edited by: Yarnold JR, Stratton MR, McmMillan TJ. Chapman and Hall; 1996:83-88.

Moody, C.A. and Laimins, L.A., 2010. Human papillomavirus oncoproteins: pathways to transformation. *Nature Reviews Cancer*, *10*(8), pp.550-560.

Moodley, I., Tathiah, N., Mubaiwa, V. and Denny, L., 2013. High uptake of Gardasil vaccine among 9-12-year-old schoolgirls participating in an HPV vaccination demonstration project in KwaZulu-Natal, South Africa. *SAMJ: South African Medical Journal*, *103*(5), pp.313-317.

Moscicki, A.B., Ellenberg, J.H., Farhat, S. and Xu, J., 2004. Persistence of human papillomavirus infection in HIV-infected and-uninfected adolescent girls: risk factors and differences, by phylogenetic type. *Journal of Infectious Diseases*, *190*(1), pp.37-45.

Munagala, R., Donà, M.G., Rai, S.N., Jenson, A.B., Bala, N., Ghim, S.J. and Gupta, R.C., 2009. Significance of multiple HPV infection in cervical cancer patients and its impact on treatment response. *International Journal of Oncology*, *34*(1), pp.263-271.

Nakagawa, F., Lodwick, R.K., Smith, C.J., Smith, R., Cambiano, V., Lundgren, J.D., Delpech, V. and Phillips, A.N., 2012. Projected life expectancy of people with HIV according to timing of diagnosis. *AIDS*, *26*(3), pp.335-343.

Nakahara, T., Peh, W.L., Doorbar, J., Lee, D. and .. Lambert, P.F., 2005. Human papillomavirus type 16 E1∧ E4 contributes to multiple facets of the papillomavirus life cycle. *Journal of Virology*, *79*(20), pp.13150-13165.

Naucler, P., da Costa, F.M., da Costa, J.L., Ljungberg, O., Bugalho, A. and Dillner, J., 2011. Human papillomavirus type-specific risk of cervical cancer in a population with high human immunodeficiency virus prevalence: case–control study. *Journal of General Virology*, *92*(12), pp.2784-2791.

Naucler, P., Ryd, W., Törnberg, S., Strand, A., Wadell, G., Elfgren, K., Rådberg, T., Strander, B., Johansson, B., Forslund, O. and Hansson, B.G., 2007. Human papillomavirus and Papanicolaou tests to screen for cervical cancer. *New England Journal of Medicine*, *357*(16), pp.1589-1597.

Nowak, R.G., Ambulos, N.P., Schumaker, L.M., Mathias, T.J., White, R.A., Troyer, J., Wells, D., Charurat, M.E., Bentzen, S.M. and Cullen, K.J., 2017. Genotyping of

high-risk anal human papillomavirus (HPV): ion torrent-next generation sequencing vs. linear array. *Virology journal*, *14*(1), p.112.

Parkin, D.M., Bray, F., Ferlay, J. and Pisani, P., 2005. Global cancer statistics, 2002. *CA: A Cancer Journal for Clinicians*, *55*(2), pp.74-108.

Parkin, D.M., Pisani, P. and Ferlay, J., 1999. Global cancer statistics. *CA: A Cancer Journal for Clinicians*, *49*(1), pp.33-64.

Parfenov, M., Pedamallu, C.S., Gehlenborg, N., Freeman, S.S., Danilova, L., Bristow, C.A., Lee, S., Hadjipanayis, A.G., Ivanova, E.V., Wilkerson, M.D. and Protopopov, A., 2014. Characterization of HPV and host genome interactions in primary head and neck cancers. *Proceedings of the National Academy of Sciences*, *111*(43), pp.15544-15549.

Peyton, C.L., Gravitt, P.E., Hunt, W.C., Hundley, R.S., Zhao, M., Apple, R.J. and Wheeler, C.M., 2001. Determinants of genital human papillomavirus detection in a US population. *Journal of Infectious Diseases*, *183*(11), pp.1554-1564.

Pitta, D.R., Sarian, L.O., Campos, E.A., Rabelo-Santos, S.H., Syrjänen, K. and Derchain, S.F., 2009. Phylogenetic classification of human papillomavirus genotypes in high-grade cervical intraepithelial neoplasia in women from a densely populated Brazilian urban region. *São Paulo Medical Journal*, *127*(3), pp.122-127.

Quek, S.C., Lim, B.K., Domingo, E., Soon, R., Park, J.S., Vu, T.N., Tay, E.H., Le, Q.T., Kim, Y.T., Vu, B.Q. and Cao, N.T., 2013. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical intraepithelial neoplasia across 5 countries in Asia. *International Journal of Gynecological Cancer*, *23*(1), pp.148-156.

Quint, W.G.V., Scholte, G., Van Doorn, L.J., Kleter, B., Smits, P.H.M. and Lindeman, J., 2001. Comparative analysis of human papillomavirus infections in cervical scrapes and biopsy specimens by general SPF10 PCR and HPV genotyping. *Journal of Pathology*, *194*(1), pp.51-58.

Radford, A.D., Chapman, D., Dixon, L., Chantrey, J., Darby, A.C. and Hall, N., 2012. Application of next-generation sequencing technologies in virology. *Journal of General Virology*, *93*(9), pp.1853-1868.

Raybould, R., Fiander, A. and Hibbitts, S., 2011. Human papillomavirus integration and its role in cervical malignant progression. *The Open Clinical Cancer Journal*, *5*(1).

Reusser, N.M., Downing, C., Guidry, J. and Tyring, S.K., 2015. HPV carcinomas in immunocompromised patients. *Journal of Clinical Medicine*, *4*(2), pp.260-281.

Richter, K., Becker, P., Horton, A. and Dreyer, G., 2013. Age-specific prevalence of cervical human papillomavirus infection and cytological abnormalities in women in Gauteng Province, South Africa. *SAMJ: South African Medical Journal*, 103(5), pp.313-317.

Rozendaal, L., Walboomers, J.M.M., Van Der Linden, J.C., Voorhorst, F.J., Kenemans, P., Helmerhosrt, T.J., Van Ballegooijen, M. and Meijer, C.J.L.M., 1996. PCR-based high-risk HPV test in cervical cancer screening gives objective risk assessment of women with cytomorphologically normal cervical smears. *International Journal of Cancer*, *68*(6), pp.766-769.

Samji, H., Cescon, A., Hogg, R.S., Modur, S.P., Althoff, K.N., Buchacz, K., Burchell, A.N., Cohen, M., Gebo, K.A., Gill, M.J. and Justice, A., 2013. North American AIDS

Cohort Collaboration on Research and Design (NA-ACCORD) of IeDEA. Closing the gap: increases in life expectancy among treated HIV-positive individuals in the United States and Canada. *PLoS One*, *8*(12), p.e81355.

Sapp, M., Volpers, C., Müller, M. and Streeck, R.E., 1995. Organization of the major and minor capsid proteins in human papillomavirus type 33 virus-like particles. *Journal of General Virology*, *76*(9), pp.2407-2412.

Serrano, B., Alemany, L., Tous, S., Bruni, L., Clifford, G.M., Weiss, T., Bosch, F.X. and de Sanjosé, S., 2012. Potential impact of a nine-valent vaccine in human papillomavirus related cervical disease. *Infectious Agents and Cancer*, *7*(1), p.38.

Shiels, M.S., Pfeiffer, R.M., Gail, M.H., Hall, H.I., Li, J., Chaturvedi, A.K., Bhatia, K., Uldrick, T.S., Yarchoan, R., Goedert, J.J. and Engels, E.A., 2011. Cancer burden in the HIV-infected population in the United States. *Journal of the National Cancer Institute*, *103*(9), pp.753-762.

Soohoo, M., Blas, M., Byraiah, G., Carcamo, C. and Brown, B., 2013. Cervical HPV Infection in Female Sex Workers: A Global Perspective. *Open AIDS Journal*, *7*.

South Africa human papillomavirus and related cancers, fact sheet 2016.

Southern, E.M., 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, *98*(3), pp.503IN3509-508IN5517.

Stern, P.L., Faulkner, R., Veranes, E.C. and Davidson, E.J., 2001. The role of human papillomavirus vaccines in cervical neoplasia. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *15*(5), pp.783-799.

Thomas, K.K., Hughes, J.P., Kuypers, J.M., Kiviat, N.B., Lee, S.K., Adam, D.E. and Koutsky, L.A., 2000. Concurrent and sequential acquisition of different genital human papillomavirus types. *The Journal of Infectious Diseases*, *182*(4), pp.1097-1102.

Tiggelaar, S.M., Lin, M.J., Viscidi, R.P., Ji, J. and Smith, J.S., 2012. Age-specific human papillomavirus antibody and deoxyribonucleic acid prevalence: a global review. *Journal of Adolescent Health*, *50*(2), pp.110-131.

Trottier, H., Mahmud, S., Prado, J.C.M., Sobrinho, J.S., Costa, M.C., Rohan, T.E., Villa, L.L. and Franco, E.L., 2008. Type-specific duration of human papillomavirus infection: implications for human papillomavirus screening and vaccination. *The Journal of Infectious Diseases, 197 (10),* pp. 1436-1447.

van Doorn, L.J., Molijn, A., Kleter, B., Quint, W. and Colau, B., 2006. Highly effective detection of human papillomavirus 16 and 18 DNA by a testing algorithm combining broad-spectrum and type-specific PCR. *Journal of Clinical Microbiology*, *44*(9), pp.3292-3298.

van Sighem, A., Gras, L., Reiss, P., Brinkman, K. and de Wolf, F., 2010. Life expectancy of recently diagnosed asymptomatic HIV-infected patients approaches that of uninfected individuals. *AIDS*, *24*(10), pp.1527-1535.

Venceslau, E.M., Bezerra, M.M., Lopes, A.C.M., Souza, É.V., Onofre, A.S.C., Melo, C.M.D., Jeraldo, V.D.L.S. and Onofre, F.B.D.M., 2014. HPV detection using primers MY09/MY11 and GP5+/GP6+ in patients with cytologic and/or colposcopic changes. *Brazilian Journal of Pathology and Laboratory Medicine*, *50*(4), pp.280-285.

Wada, N., Jacobson, L.P., Cohen, M., French, A., Phair, J. and Muñoz, A., 2013. Cause-specific life expectancies after 35 years of age for human immunodeficiency

syndrome-infected and human immunodeficiency syndrome-negative individuals followed simultaneously in long-term cohort studies, 1984–2008. *American Journal of Epidemiology*, *177*(2), pp.116-125.

Walboomers, J.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., Snijders, P.J., Peto, J., Meijer, C.J.L.M. and Munoz, N., 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *Journal of Pathology*, *189*(1), pp.12-19.

Winer, R.L., Kiviat, N.B., Hughes, J.P., Adam, D.E., Lee, S.K., Kuypers, J.M. and Koutsky, L.A., 2005. Development and duration of human papillomavirus lesions, after initial infection. *Journal of Infectious Diseases*, *191*(5), pp.731-738.

WHO-ICO. (2010). Human papillomavirus and Relate Cancers: Summary Report Update, Information Centre of HPV and Cervical Cancer, South Africa. http;//www.who.int/hpvcentre.

WHO. (2010). Human papillomavirus vaccines*: WHO position paper, Weekly Epidemiological Record 2014*. **43(89):** 465-92.

Zhang, R., Shi, T.Y., Ren, Y., Lu, H., Wei, Z.H., Hou, W.J., Zhang, M. and Xu, C., 2013. Risk factors for human papillomavirus infection in Shanghai suburbs: a population-based study with 10,000 women. *Journal of Clinical Virology*, *58*(1), pp.144-148.

Zhou, X.H., Shi, Y.F., Wang, L.J., Liu, M. and Li, F., 2017. Distribution Characteristics of Human Papillomavirus Infection: A Study Based on Data from Physical Examination. *Asian Pacific Journal of Cancer Prevention: APJCP*, *18*(7), p.1875.