Genetic diversity in *L1* gene of human papillomavirus variants in individuals with cervical cancer with and without human immunodeficiency virus in Botswana and Kenya

Leabaneng Tawe (tleabaneng@yahoo.com)
University of Botswana
Wonderful T. Choga
Botswana Harvard AIDS Institute Partnership
Giacomo M. Paganotti
Botswana-University of Pennsylvania Partnership
Ontlametse T. Bareng
University of Botswana
Tlhalefo D. Ntereke
Botswana-University of Pennsylvania Partnership
Pleasure Ramatlho
University of Botswana
Doreen Ditshwanelo
Botswana Harvard AIDS Institute Partnership
Simani Gaseitsiwe
Botswana Harvard AIDS Institute Partnership
Ishmael Kasvosve
University of Botswana
Doreen Ramogola-Masire
Department of Obstetrics and Gynaecology, Faculty of Medicine
Omenge E. Orang'o
Moi University
Erle Robertson
University of Pennsylvania
Nicola Zetola
Botswana-University of Pennsylvania Partnership
Sikhulile Moyo
Botswana Harvard AIDS Institute Partnership
Surbhi Grover
Botswana-University of Pennsylvania Partnership

Aaron C. Ermel
Indiana University School of Medicine

Research Article

Keywords: Botswana, Kenya, cervical cancer, human papillomavirus, HPV variants phylogenetic analysis, HIV, L1 gene, HIV co-infection

Posted Date: July 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-668829/v1

Abstract

Background

The variation of human papillomavirus (HPV) genotypes shapes the risks of cervical cancer and these variations are not well defined in Africa. Nucleotide changes within the *L1* gene, nucleotide variability, and phylogeny were explored in relation to HIV in samples from Botswana and Kenya.

Methods

A total of 98 HPV-positive cervical samples were sequenced to identify different HPV variants. Phylogenetic inferences were used to determine HPV genotypes and investigate the clustering of sequences between women living with HIV (WLWHIV) and -women not living with HIV (WNLWHIV).

Results

Out of 98 generated sequences, 83.7% (82/98) participants had high-risk(HR) HPV genotypes while 16.3% (16/98) had low-risk (LR) HPV genotypes. Among participants with HR-HPV genotypes, 47.6% (39/82) were coinfected with HIV. The prevalence of HR-HPV genotypes was statistically higher in the Botswana population compared to Kenya (p-value < 0.001). Multiple amino acid mutations were identified in both countries. Genetic diversity differed considerably among WLWHIV and WNLWHIV. The mean pairwise distances between HPV-16 between HIV and HIV/HPV as well as for HPV-18 were statistically significant. Six (6) new deleterious mutations were identified in the HPV genotypes based on the sequencing of the L1 region, HPV-16 (L441P, S343P), HPV-18 (S424P), HPV-45 (Q366H, Y365F), and HPV-84 (F458L). The majority of the patients with these mutations were co-infected with HIV.

Conclusions

Genomic diversity and different genomic variants of HPV sequences were demonstrated. Candidate novel mutations within the L1 gene were identified in both countries which can be further investigated using functional assays.

Background

Cervical cancer continues to be a major public health problem particularly in less-resourced countries with an estimated 570 000 cases diagnosed, and 311 000 mortality rates in 2018 [1]. It is one of the most common cancers in women living with human immunodeficiency virus (HIV) and presents a significant public health threat to women especially on the African continent. In sub-Saharan Africa, human papillomavirus (HPV)-associated cervical cancer is an important cause of morbidity and mortality [2]. Cervical cancer has been recognized as a rare outcome of a sexually transmitted infection, and the

etiology is limited to a few HPV genotypes. HPV infections can be facilitated by co-infection with HIV [3]. To date, over 100 HPV genotypes have been identified and classified according to their oncogenic potential, they are divided into high risk (HR), possible or probable high risk and low risk (LR) HPV genotypes, depending on their association with the development of cancer. For decades, 14 highly carcinogenic HPV genotypes (-16, -18, -31, -33, -35, -39, -45, -51, -52, -56, -58, -59, -66 and − 68) have been recognized as the causative agent of cervical cancer [4, 5]. However, HPV genotypes − 16 and − 18 have been demonstrated and identified as the predominant oncogenic genotypes of all cervical cancer cases worldwide [6]. Despite having the highest burden of risk factors associated with HPV infection, persistence, and progression to cervical cancer [7], comprehensive data on genotypes and profiles of mutations associated with different HPV risk groups such as HIV/HPV co-infected in Africa are lacking.

The HPV particle is 50-60nm in diameter, and its surface consists of 72 capsomeres [8]. Wrapped inside the capsid proteins is the HPV genome consisting of double-stranded DNA measuring approximately 7-8kbp. The HPV genome may be divided into three regions, early (E: *E1, E2, E3, E4, E5, E6, E7*, and *E8* genes), late (L: *L1* and *L2* genes), and (non-coding) long control region. The *E* region is crucial for HPV replication, transcription, translation, and transformation. The L region (~ 2500 bp) encodes the two structural proteins (*L1* and *L2*) and contains functional regulators for HPV replication and transcription [9]. The *L1* gene in the HPV genome is highly conserved and has been shown to provide the basis for HPV genotyping. An HPV 'type' is designated when the nucleotide sequence of the *L1* open reading frames (ORF) from the cloned viral genome is more than 10% dissimilar to all known types [10]. Molecular epidemiological studies assessing the phylogenetic association of HPVs based on oncogenic risk and supporting specific biological and pathological traits distinctive to HPV genera, species and types are still limited. Previous studies in Africa have explored the frequency and distribution of the HPV genotypes in HIV-infected women [11], with consistent results about the influence of HIV on HPV genotype distribution. Several studies in Africa have also explored the genomic diversity of HPV variants [12–17]. HPV genotypes − 16, -18, -45, and − 58 were observed among most cervical samples studied from Botswana and Kenya [2, 16, 25, 18–24, 26–28].

The objective of this study was to use Sanger genotyping protocol to detect and genotype HPV isolates in invasive cervical cancers specimens of cervical cancer patients' living- and not living- with HIV obtained from Botswana and Kenya. Apart from determining the prevalence and distribution of HPV genotypes (LR and HR) among the two groups (HPV versus HIV/HPV-coinfected), the mutation profiles (nucleotide) and amino acid changes) within *L1* region of HPV genotypes were used to determine the mean pairwise distances and assessing for any signature mutations associated with immune pressure among the HIV/HPV-coinfected patients. Phylogenetic analyses were used to (i) assign HPV genotypes, and (ii) determine clustering of sequences within and between the two countries.

Methods

Study design and population

Formalin-fixed paraffin-embedded (FFPE) tissue specimens from women living with HIV (WLWHIV) and -women not living with HIV (WNLWHIV), diagnosed with invasive cervical cancer previously typed using the Abbot real-time polymerase chain reaction (PCR) and Linear Array HPV Genotyping Test, LA-HPV (Roche Applied Sciences, Indianapolis, IN) from prior retrospective cross-sectional studies [21,23] from Botswana and Kenya were utilized. We proposed to describe the sequence variation of samples with single HPV infections for the two HR-HPV genotypes (HPV -16 and -18) reported in both countries within the ~450 base pairs (bp) region amplified by these HPV genotyping systems. Demographic data including age, cancer stage, country of origin and HIV status were obtained for samples from Botswana, However, only HIV status was available for Kenya samples.

DNA extraction

DNA extraction was performed from the tissue samples archived using a previously established protocol [29]. The extracted DNA was stored at -80°C prior to analysis. For the samples from Botswana, the presence of HR-HPV DNA was detected using Abbot real-time PCR (Abbot Molecular Inc., Chicago) [23]. While specimens from Kenyan women, HR-HPV detection and genotyping had been previously performed using the LA-HPV [21].

Polymerase chain reaction (PCR) and sequencing

To isolate a region of *L1* targeted by the LA-HPV and the Abbott real-time PCR method, conventional PCR for HPV DNA viral amplification of a ~450bp HPV-specific segment from the *L1* gene covering nucleotide positions (5722- 6162) numbered according to NC001526 HPV -16 reference genome and was performed using the primer MY09 5'-CGTCCMARRGGAWACTGATC-3' and the reverse primer MY11 5'-GCMCAGGGWCATAAYAATGG-3'. Five microliters of DNA were added to 15 µL of reaction mix containing 1× PCR buffer, 0.2 mM dNTPs, 4 mM $MgCl_2$, 0.3 µM of each primer, and 2 U/µL of Platinum Taq DNA Polymerase, High fidelity (Invitrogen, USA). PCR products were subjected to electrophoresis in 4% agarose (Applichem) using 1× TBE buffer (Applichem) and visualized under UV light.

The thermocycling conditions were denaturing at 96°C for 10 seconds, annealing at 50°C for 5 seconds, and final extension at 60°C for 4 minutes for 25 cycles. PCR products were then purified using commercially available SureClean Plus (Bioline) and the purified products were sequenced directly via automated sequencing using two overlapping PCR primers (both forward and reverse). The BigDye Terminator v3.0 kit (Applied Biosystems; Foster City, CA, USA) was used for sequencing using the automated Sequencer (ABI PRISM 3130xl; Applied Biosystems).

Sequencing and sequence editing

The generated chromatographs were quality assessed using Sequencher v5.0 software (Gene Codes Corp., Ann Arbor, MI, USA) followed by generation of consensus sequences in FASTA file format, which were used, for downstream analysis. An online *L1* Taxonomy Tool Analysis Results was utilized to check for similarity with the known HPV reference sequences. Additionally, the BLAST tool was used to

determine the subset of sequences with high E^0 value that can be included in the phylogenetic tree. Finally, referenced alignments per each genotype were performed using AliView version 1.26 and were used for phylogenetic analysed while the translated alignment was used for mutation analysis. We developed a specific workflow to answer the study objectives (Figure 1).

Phylogenetic analysis

Phylogenetic inference was performed using the Bayesian Markov chain Monte Carlo (MCMC) approach implemented in the Bayesian Evolutionary Analysis with the Sampling Trees software (BEAST v1.10.2) with an uncorrelated log-normal relaxed molecular clock, General Time Reversible substitution model, and gamma site heterogeneity. The MCMC was set at a chain length of 500,000,000 with parameters logged every 10,000. The tree was visualized in FigTree v1.4.3 after a 10% burn-in using Tree Annotator v1.8.4. Posterior probabilities 0.90 and above were noted as statistically significant. The sequences generated in this study are available in GenBank under accession numbers (awaiting accession numbers)

HPV-specific mutation analysis

To assess for any known mutations or signatures amino acids that may be associated with immune pressure among HPV patients in Botswana and Kenya, 95 aligned sequences with high coverage were included in the analysis (N_1 dataset). Here, the top 100 sequences after sorting by E value and covering the full *L1* region were obtained from the NCBI blast search and compared to the N_1 dataset. Comparisons were done at the ***aa*** level to exclude synonymous polymorphisms. Genetic diversity and any signature mutations associated with HIV/HPV co-infection were investigated by comparing the HPV sequences isolated from patients with HPV mono-infection versus those with HPV mono-infection. Mutations that had not been noted previously in the literature before were termed “candidate novel mutations”. The candidate novel mutations were further assessed using PROVEAN [30] and SNP2 [31] to determine their impact at the gene level. Briefly, PROVEAN classifies mutations with a negative impact on protein biological function as deleterious and SNAP2 uses ‘effect’ or ‘neutral’ to indicate the presence or absence of change in protein function caused by a mutation as discussed elsewhere.

Statistical analysis

Raw data were collected, processed and coded using Excel. Study demographics were presented as percentages for categorical variables and compared among participants with HR-HPV and LR-HPV HPV genotype using the Chi-square test. Wilcoxon rank-sum test was utilised to compare continuous variables. Differences among the prevalence of HPV genotypes in Botswana and Kenya was assessed using a comparison of proportion test. All the statistical analysis was done using Stata version 15 (Stata Corp, College Station, TX, USA) and p-value > 0.05 was considered not significant.

Results

Population characteristics

The present study sought to look at the variation within the *L1* region of the HPV genome in 98 samples. All participants were female and a subpopulation was from Botswana (n = 72) and another from Kenya (n = 26). The distribution of HPV genotypes stratified by age and country are shown in Figure 2. Out of 98 generated sequences, 83.7% (82/98) participants had HR-HPV genotypes while 16.3% (16/98) had LR-HPV genotypes. However, in some patients, we identified other additional HPV genotypes which were previously missed when the same samples were genotyped using Abbott and LA-HPV. Among participants with HR-HPV genotypes, 51.2% (42/82) were coinfected with HIV (HIV/HPV), while 47.5% (39/82) were from Botswana and 3.7% (3/82) were from Kenya based on MY09/11 Sanger sequencing-based method. We could not assess the cancer stage of HPV sequences isolated from Kenyan participants because their demographics and cancer stage information at diagnosis were not available. The most predominant cancer stage among participants from Botswana was stage 2. Amongst individuals with LR-HPV genotypes, 12.5% were infected with HIV, cancer stage 3 was the most predominant cancer stage. The prevalence of HR-HPV genotypes was statistically higher in the Botswana population compared to Kenya (p-value < 0.001). We did not record any statistical significance in other variables among HR and LR- HPV genotypes among the Botswana population.

Table 1. Baseline demographics for participants

	Total n=98	**HR-HPV n=82**	**LR-HPV n=16**	***P*-value**
ledian Age in years Median (IQR)	50 (42-54)	50 (42-61)	51.5 (47.5-57)	*0.67*[a]
IV Status, n (%)				*0.22*[b]
IV negative	31 (31.6)	27 (32.9)	4 (25)	
IV positive	41 (41.8)	39 (47.6)	2 (12.5)	
ountry, n (%)				*<0.001*[b]
otswana	72 (73.5)	66 (80.5)	6 (37.5)	
enya	26 (26.5)	16 (19.5)	10 (62.5)	
ancer Stage, n (%) Botswana only				*0.69*[b]
tage 1	6 (6.1)	6 (7.3)	0 (0.0)	
tage 2	21 (21.4)	18(22.0)	3 (18.8)	
tage 3	15 (15.3)	14 (17.0)	1 (6.3)	
tage 4	1 (1.0)	1 (1.2)	0 (0.0)	

[a] *P*-value was calculated using the Rank sum test

[b] *P*-values were obtained by chi-square test

HIV, human immunodeficiency virus; HPV, human papillomavirus; HR, high risk; LR, low risk; IQR, interquartile range

HPV genotypes

Human papillomavirus genotypes were identified using both phylogenetic tree (Figure 3) and online *L1* Taxonomy Tool Analysis. After quality control (QC), 3 sequences were excluded for mutation and phylogenetic analyses because sequences were too short (>450bp). In total, 19 HPV genotypes were determined, -6, -11, -16, -18, -35, -39, -42, -45, -53, -54, -55, -56, -58, -59, -73, -81, -89, -122, and -159 (Figure 2). Out of 19 different genotypes found in Botswana and Kenya populations, five showed statistically significant differences in their frequency (p-value < 0.05) between countries (Figure 2). Thus, HPV genotypes -6, -54 and -73 were higher in Kenyan population while genotypes -16 and -18 were recorded higher in Botswana population. The WLWHIV had more HPV genotypes compared to WNLWHIV even though the difference was not statistically significant (p-value > 0.05).

Phylogenetic analysis

Phylogenetic analysis included all the 95 HPV sequences obtained in this study that were adequate for phylogeny (>400bp sequence length). All the studied HPV sequences clustered with reference and had posterior probabilities > 90%, and could be used to assign the genotypes with confidence. Phylogenetic inference with maximum likelihood and MCMC methods showed that there were no isolated clusters among HPV sequences from both countries. This was also true when trees for HPV-16 (n=29) and HPV-18 (n=12) sequences from this study and the respective GenBank references for the *L1* gene region were constructed (Figure 4). However, there was a general increase in nucleotide genetic diversity among the HPV sequences isolated from WLWHIH as opposed to WNLWHIV (Figure 4) as shown by multiple branches within the phylogenetic tree. The overall mean pairwise distances of HPV-16 and HPV-18 sequences isolated from WLWHIV versus WNLWHIV were not statistically significant.

Mutations Analysis

Nucleotide Diversity

In total, 68 nucleotide base substitutions were detected using reference sequences for each genotype. Table 1 summarises the different polymorphisms detected per genotype and only genotypes with counts greater than 1 per strata were considered. Although there were more HPV types among WNLWHIV, most of the HPV genotypes among WLWHIV were HR-HPV genotypes.

Mutations at the Amino Acid level

The present study defined mutations as amino acid changes that differed from that of reference sequence and candidate escape mutations were those that have not yet been reported in 1000 most similar sequences. Six (6) new variants were identified based on the sequencing of the *L1* region, HPV-16 (L441P, S343P), HPV-18 (S424P), HPV-45 (Q366H, Y365F), belonging to the HPV HR-HPV group and HPV-84 (F458L) belonging to the LR-HPV group.

Table 2. Novel genetic mutations found in the 5′ and 3′-ends of HPV-16, -18, -45, -84 -*L1* regions.

HPV genotype	Mutations	HIV status	Cancer Stage	Count of Sequences in GRS	Deleterious
16	L441P	Positive	2	-	√
	S343P	Positive	n/a	-	√
18	S424P	Positive	n/a	-	√
45	Q366H	Positive	3	51	√
84	Y365F	Negative	2	-	√
	F458L	Negative	n/a	-	√

GRS, gene recruitment sequence, HIV, human immunodeficiency virus; HPV, human papillomavirus; n/a, not applicate

Discussion

Given the uniquely large and diverse collection of HPV-genotyped cervical samples worldwide, we were able to evaluate the genetic diversity within HPV genotypes and report on the geographic distribution of HPV variants, as well as measure their association with cervical cancer in Botswana and Kenya. To date, several studies have reported varying prevalence of HPV in the general population and in cervical cancer patients. Though HPV is the main causative agent for cervical cancer, its prevalence and distribution vary in different geographical regions of the world. Several studies have shown HPV prevalence in Botswana and Kenya [2, 16, 25, 18–24, 26–28]. However, the numbers of studies investigating the variants within HPV genotypes are limited.

In the present study, the molecular characterization of HPV variants within the MY09/MY11 *L1* genomic region was performed in 98 sequences from Botswana and Kenya. By sequencing the *L1* region of 98 HPV positive cervical samples, we were able to confirm the majority of the previously reported genotypes [18–23]. However, other HPV genotypes were identified through sequencing that were not previously detected in both countries (See supplementary Table 2).These HPV genotypes were HPV-159 and HPV-122 in Botswana and HPV genotypes – 21, -44, -54, -56, -6, -81, -84 and – 89. This is likely due to the fact that the MY09/11 primers are universal primers and sequencing is not limited by the presence of HPV type-specific probes as in the LA-HPV and Abbott realtime PCR methods. Again, detection from FFPE samples could have been suboptimal due to nonuniform coverage of the tissue. In some cases, low-level of HPV replication may be missed because of thresholds standards. Assessing the quality of extracted DNA aids in selecting the optimal sequencing approach, and the choice of both DNA extraction and library preparation approaches can impact the performance of archival tissue in sequencing. DNA extracted from FFPE specimens presents degradation due to specimen processing [32] such as nucleic acid fragmentation, DNA crosslinks, a basic site leading to localized DNA denaturation and strand breaks, and deamination which impede downstream sequencing analysis [33]. Overall, the frequency of the genotypes identified in both countries are the same as previously reported, having HPV-16 as the most frequently detected genotype followed by HPV-18 and the rest. The sequencing analysis illustrated that multiple variants were identified in both countries. Variants of HPV that changed the protein sequence of the capsid protein encoded by *L1* were found. These observations, in light of the slow evolution of HPV at the population level, suggest that strong selection pressures are at play in each infection cycle. We analyzed the mutation frequency in the panel for HPV genotypes in WLWHIV and WNLWHIV and found mutations within HPV genotypes. We could not assess the clinical relevance (i.e., association with HIV infection or cancer stage at presentation) of the detected variants due to small sample size.

The most striking finding in this report is the high proportion of mutations in the HPV-genotypes (all HPV-related cases, including both those arising among WLWHIV and those arising among WNLWHIV). Several nucleotide sequence variations were found in the *L1* region of the HPV genotypes (Supplementary Table 1). In this study, 6 novel nucleotide variations, which were previously unreported in the literature, were found in HPV-16, HPV-18, HPV-45 and HPV-84 (Table 2). However, two of them were found in only one sample (Y365F and F458L) and may have occurred by PCR amplification. These variations could be related to the early promoter activation of HPV and may play a crucial role in the transcriptional modulation of the HPV *L1* oncogenes via the promoter. Variants identified based on the sequencing of the *L1* region are; HPV-16 (L441P, S343P), HPV-18 (S424P), HPV-45 (Q366H, Y365F), belonged to the HPV HR-HPV group and HPV-84 (F458L) for the LR-HPV group. These alterations are described for the first time, and functional implications resulting from this variation need further analyses. To determine whether the synonymous mutations that we described have any impact in the protein expression, additional analysis is required. Changes in the *L1* region of the HPV genome may be important for discriminating the infectious potential of different variants, as well as in defining epitopes relevant to vaccine design. Some previous studies investigating HPV-16 full-length sequences in cervical specimens have shown that the contiguous deletions identified to be highly associated with cancer are suggestive of

a pattern of HPV integration [34, 35]. The findings of this study indicate that there could be variants of HPV circulating within sub-Saharan Afrcia. Further studies are needed to confirm the presence of new HPV variants and genotypes and to understand the evolution of HPV isolates in Botswana and Kenya by analyzing the complete HPV genome or different regions of HPV genes such as *L2*, LCR, *E6* and *E7*.

Our work showed genetic variability of *L1*, making it essential to take into account the HPV variants lineages or population stratification when developing vaccines. The limitation of this study was the small sample size for samples from Botswana and Kenya; we could not determine whether the synonymous mutations that are described have any impact on the protein expression. We attempted to investigate HPV full-length sequences in FFPE cervical cancer specimens, but were unable to do so because of highly fragmented samples. Further study is required to determine whether variants represent a higher risk for cervical cancer. Additionally, functional studies regarding HPV polymorphisms across the HPV genome variants should be conducted to explore the biological evidence of carcinogenicity.

Conclusion

We demonstrated the genomic diversity of HPV sequences and phylogenesis of HPV genotypes giving important information as 6 new *L1* single nucleotide changes were identified. However, it was not possible to correlate disease severity with any particular variant. The results illustrated that the distribution of HPV genotypes differs between two countries and multiple variants were identified in both countries. Further studies with complete sequencing of HPV genomes from large population-based and case-control studies of cervical pre-cancer and cancer are required to understand viral carcinogenesis and possibly to improve preventive and therapeutic strategies in the future.

Abbreviations

HIV: human immunodeficiency virus; HPV: human papillomavirus; HR: high risk; LR: Low risk; WLWHIV: women living with HIV; WNLWHIV: women not living with HIV; FFPE: Formalin-fixed paraffin-embedded; LA-HPV: Linear Array HPV Genotyping Test; MCMC: Markov chain Monte Carlo; PCR: polymerase chain reaction; bp: base pair

Declarations

Ethics approval and consent to participate

The research was approved by Institutional Review Board (IRB) at the University of Botswana (RES/IRB/1636), the Human Resource Development Council at the Botswana Ministry of Health and Wellness and the University of Pennsylvania's IRB (HPDME 13/18/1 X1). Approval for the study was also granted by the Institutional Review Board at Indiana University School of Medicine, and by the Institutional Research and Ethics Committee at Moi University Teaching and Referral Hospital (IREC/2011/174). Material Transfer Agreement (MTA) was obtained from Institutional Review

Board at Indiana University School of Medicine and the Institutional Review Board (IRB) at the University of Botswana. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analyzed during the current study are available

from the corresponding author on reasonable request.

Competing interests

There are no relevant conflicts of interest for the authors of the study.

Funding

This work was supported by the Sub-Saharan African Collaborative HIV and Cancer Consortia-U54 [grant # 1U54 CA190158-01], the American Cancer Society International Fellowships for Beginning Investigators (ACSBI), Conquer Cancer Foundation Young Investigator Award, Mentored Patient-Oriented Career Research Development Award (1-K08CA230170-01A1) and Penn Center for AIDS Research [grant # 5-P30-AI- 045008-17]. This work was also supported by the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a DELTAS Africa Initiative [grant # DEL-15-006]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant # 107752/Z/15/Z]. The funding bodies had no role in the design of the study, in the collection, analysis, and interpretation of data, or in writing the manuscript.

Authors' contributions

LT: Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Methodology, Supervision, Writing - original draft, Writing - review & editing. WTC: Data curation, Methodology, Validation, Formal analysis, Writing - original draft, Writing - review & editing. GMP: Data curation, Methodology, Validation. Writing - original draft, Writing - review & editing. OTB: Investigation, Formal analysis. TDN: Data curation, Investigation. PR: Investigation. DD: Investigation. SG: Investigation, Supervision. IK: Investigation, Supervision. DR: Investigation, Supervision. OEO: Investigation, Supervision. ER: Investigation, Supervision. NZ: Investigation, Supervision. SM: Conceptualization, Formal analysis, Methodology, Investigation, Supervision, Validation, Writing - review & editing. SG: Conceptualization, Investigation, Supervision, Validation, Funding acquisition, Writing - review & editing. ACE: Conceptualization, Investigation, Supervision, Validation, Funding acquisition, Writing - review & editing. All authors have read and approved the manuscript.

Acknowledgements

We would like to acknowledge the contributions of Ms Dorcas Maruapula, and Mr Patrick Mokgethi from Botswana Harvard AIDS Institute, who provided insight and expertise that greatly assisted the completion of the study.

References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer. 2019; 144(8):1941–53. doi: 10.1002/ijc.31937.
2. De Vuyst H, Chung MH, Baussano I, Mugo NR, Tenet V, van Kemenade FJ, et al. Comparison of HPV DNA testing in cervical exfoliated cells and tissue biopsies among HIV-positive women in Kenya. Int J Cancer. 2013; 133(6):1441–6. doi: 10.1002/ijc.28131. Epub 2013 Mar 16.
3. Badial RM, Dias MC, Stuqui B, Melli PPDS, Quintana SM, Bonfim CMD, et al. Detection and genotyping of human papillomavirus (HPV) in HIV-infected women and its relationship with HPV/HIV co-infection. Medicine (Baltimore). 2018; 97(14):e9545. doi: 10.1097/MD.0000000000009545.
4. Burd EM. Human papillomavirus and cervical cancer. Clin Microbiol Rev. 2003; 16(1):1–17. doi: 10.1128/CMR.16.1.1-17.2003.
5. Muñoz N, Bosch FX, de Sanjosé S, Herrero R, Castellsagué X, Shah KV, et al. International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med. 2003; 348(6):518–27. doi: 10.1056/NEJMoa021641.
6. de Sanjose S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, Tous S, et al. Retrospective International Survey and HPV Time Trends Study Group. Human papillomavirus

genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. Lancet Oncol. 2010; 11(11):1048–56. doi: 10.1016/S1470-2045(10)70230-8.

7. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011; 61(2):69–90. doi: 10.3322/caac.20107.
8. Chao KC, Wang PH, Yen MS, Chang CC. Detection of HPV infection by analyzing the changes in structure of peripheral blood lymphocytes specifically induced by HPV E7 antigen. Eur J Gynaecol Oncol. 2003; 24(1):30–2.
9. Morshed K, Polz-Gruszka D, Szymański M, Polz-Dacewicz M. Human Papillomavirus (HPV) - structure, epidemiology and pathogenesis. Otolaryngol Pol. 2014; 68(5):213–9. doi: 10.1016/j.otpol.2014.06.001. Epub 2014 Jun 27.
10. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. Virology. 2004; 324(1):17–27. doi: 10.1016/j.virol.2004.03.033.
11. Clifford GM, de Vuyst H, Tenet V, Plummer M, Tully S, Franceschi S. Effect of HIV Infection on Human Papillomavirus Types Causing Invasive Cervical Cancer in Africa. J Acquir Immune Defic Syndr. 2016; 73(3):332–9. doi: 10.1097/QAI.0000000000001113.
12. Tornesello ML, Buonaguro FM, Meglio A, Buonaguro L, Beth-Giraldo E, Giraldo G. Sequence variations and viral genomic state of human papillomavirus type 16 in penile carcinomas from Ugandan patients. J Gen Virol. 1997; 78:2199–208. doi: 10.1099/0022-1317-78-9-2199.
13. Calleja-Macias IE, Kalantari M, Huh J, Ortiz-Lopez R, Rojas-Martinez A, Gonzalez-Guerrero JF, et al. Genomic diversity of human papillomavirus-16, 18, 31, and 35 isolates in a Mexican population and relationship to European, African, and Native American variants. Virology. 2004; 319(2):315–23. doi: 10.1016/j.virol.2003.11.009.
14. Jendoubi-Ferchichi M, Satouri L, Ghoul F, Malek-Mellouli M, Derbel AM, Makni MK, Reziga H, et al.. Phylogeny and Classification of Human Papillomavirus (HPV)16 and HPV18 Variants Based on E6 and L1 genes in Tunisian Women with Cervical Lesions. Asian Pac J Cancer Prev. 2018; 19(12):3361–3366. doi: 10.31557/APJCP.2018.19.12.3361.
15. Fitzpatrick MB, Hahn Z, Mandishora RSD, Dao J, Weber J, Huang C, et al. Whole-Genome Analysis of Cervical Human Papillomavirus Type 35 from rural Zimbabwean Women. Sci Rep. 2020; 10(1):7001. doi: 10.1038/s41598-020-63882-z.
16. Omire A, Budambula NLM, Kirumbi L, Langat H, Kerosi D, Ochieng W, et al. Cervical Dysplasia, Infection, and Phylogeny of Human Papillomavirus in HIV-Infected and HIV-Uninfected Women at a Reproductive Health Clinic in Nairobi, Kenya. Biomed Res Int. 2020; 2020:4945608. doi: 10.1155/2020/4945608.
17. Pinheiro M, Gage JC, Clifford GM, Demarco M, Cheung LC, Chen Z, et al. Association of HPV35 with cervical carcinogenesis among women of African ancestry: Evidence of viral-host interaction with implications for disease intervention. Int J Cancer. 2020; 147(10):2677–86. doi: 10.1002/ijc.33033.
18. Ramogola-Masire D, McGrath CM, Barnhart KT, Friedman HM, Zetola NM. Subtype distribution of human papillomavirus in HIV-infected women with cervical intraepithelial neoplasia stages 2 and 3

in Botswana. Int J Gynecol Pathol. 2011; 30(6):591–6. doi: 10.1097/PGP.0b013e31821bf2a6.

19. Macleod IJ, O'Donnell B, Moyo S, Lockman S, Shapiro RL, Kayembe M, et al.. Prevalence of human papillomavirus genotypes and associated cervical squamous intraepithelial lesions in HIV-infected women in Botswana. J Med Virol. 2011; 83(10):1689–95. doi: 10.1002/jmv.22178.
20. Ermel A, Ramogola-Masire D, Zetola N, Tong Y, Qadadri B, Azar MM, et al. Invasive cervical cancers from women living in the United States or Botswana: differences in human papillomavirus type distribution. Infect Agent Cancer. 2014; 9:22. doi: 10.1186/1750-9378-9-22.
21. Ermel A, Qadadri B, Tong Y, Orang'o O, Macharia B, Ramogola-Masire D, Zetola NM, Brown DR. Invasive cervical cancers in the United States, Botswana and Kenya: HPV type distribution and health policy implications. Infect Agent Cancer. 2016; 11:56. doi: 10.1186/s13027-016-0102-9.
22. Rantshabeng P, Kasvosve I, Ndlovu A, Gaseitsiwe S, Moyo S. Prevalence of high-risk human papilloma virus in women with high-grade squamous cell intraepithelial lesions in Botswana using Abbott RealTime HPV assay. PLoS One. 2019; 14(1):e0211260. doi: 10.1371/journal.pone.0211260.
23. Tawe L, MacDuffie E, Narasimhamurthy M, Wang Q, Gaseitsiwe S, Moyo S, et al. Human papillomavirus genotypes in women with invasive cervical cancer with and without human immunodeficiency virus infection in Botswana. Int J Cancer. 2020; 146(6):1667–1673. doi: 10.1002/ijc.32581. Epub 2019 Aug 2.
24. Karani LW, Musyoki S, Orina R, Nyamache AK, Khayeka-Wandabwa C, Nyagaka B. Human papillomavirus genotype profiles and cytological grades interlinkages in coinfection with HIV. Pan Afr Med J. 2020;35:67. doi: 10.11604/pamj.2020.35.67.21539.
25. Guthrie BL, Rositch AF, Cooper JA, Farquhar C, Bosire R, Choi R, et al. Human papillomavirus and abnormal cervical lesions among HIV-infected women in HIV-discordant couples from Kenya. Sex Transm Infect. 2020;96(6):457–463. doi: 10.1136/sextrans-2019-054052.
26. Menon S, van den Broeck D, Rossi R, Ogbe E, Mabeya H. Multiple HPV infections in female sex workers in Western Kenya: implications for prophylactic vaccines within this sub population. Infect Agent Cancer. 2017; 12:2. doi: 10.1186/s13027-016-0114-5.
27. Orang'o EO, Emont JP, Ermel AC, Liu T, Omodi V, Tong Y, et al. Detection of types of HPV among HIV-infected and HIV-uninfected Kenyan women undergoing cryotherapy or loop electrosurgical excision procedure. Int J Gynaecol Obstet. 2020;151(2):279–86. doi: 10.1002/ijgo.13330.
28. Luchters SM, Vanden Broeck D, Chersich MF, Nel A, Delva W, Mandaliya K, et al. Association of HIV infection with distribution and viral load of HPV types in Kenya: a survey with 820 female sex workers. BMC Infect Dis. 2010;10:18. doi: 10.1186/1471-2334-10-18.
29. Tawe L, Grover S, Narasimhamurthy M, Moyo S, Gaseitsiwe S, Kasvosve I, et al. Molecular detection of human papillomavirus (HPV) in highly fragmented DNA from cervical cancer biopsies using double-nested PCR. MethodsX. 2018; 5:569–578. doi: 10.1016/j.mex.2018.05.018.
30. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One.;7(10):e46688. doi: 10.1371/journal.pone.0046688.

31. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. BMC Genomics. 2015;16 Suppl 8(Suppl 8):S1. doi: 10.1186/1471-2164-16-S8-S1.
32. Bass BP, Engel KB, Greytak SR, Moore HM. A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen? Arch Pathol Lab Med. 2014;138(11):1520–30. doi: 10.5858/arpa.2013-0691-RA.
33. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin Chem. 2015;61(1):64–71. doi: 10.1373/clinchem.2014.223040.
34. Cullen M, Boland JF, Schiffman M, Zhang X, Wentzensen N, Yang Q, et al. Deep sequencing of HPV16 genomes: A new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection. Papillomavirus Res. 2015;1:3–11. doi: 10.1016/j.pvr.2015.05.004.
35. Sun M, Gao L, Liu Y, Zhao Y, Wang X, Pan Y, et al. Whole genome sequencing and evolutionary analysis of human papillomavirus type 16 in central China. PLoS One. 2012;7(5):e36577. doi: 10.1371/journal.pone.0036577.
36. Mane A, Patil L, Limaye S, Nirmalkar A, Kulkarni-Kale U. Characterization of major capsid protein (L1) variants of Human papillomavirus type 16 by cervical neoplastic status in Indian women: Phylogenetic and functional analysis. J Med Virol. 2020;92(8):1303–1308. doi: 10.1002/jmv.25675.
37. Ahmed HG, Bensumaidea SH, Alshammari FD, Alenazi FSH, ALmutlaq BA, Alturkstani MZ, et al. Prevalence of Human Papillomavirus subtypes 16 and 18 among Yemeni Patients with Cervical Cancer. Asian Pac J Cancer Prev. 2017;18(6):1543–1548. doi: 10.22034/APJCP.2017.18.6.1543.
38. Ntova CK, Kottaridi C, Chranioti A, Spathis A, Kassanos D, Paraskevaidis E, et al. Genetic variability and phylogeny of high risk HPV type 16, 18, 31, 33 and 45 L1 gene in Greek women. Int J Mol Sci. 2012;13(1):1–17. doi: 10.3390/ijms13010001.
39. Pande S, Jain N, Prusty BK, Bhambhani S, Gupta S, Sharma R, et al. Human papillomavirus type 16 variant analysis of E6, E7, and L1 genes and long control region in biopsy samples from cervical cancer patients in north India. J Clin Microbiol. 2008;46(3):1060–6. doi: 10.1128/JCM.02202-07.
40. Gagnon S, Hankins C, Tremblay C, Forest P, Pourreaux K, Coutlée F; Canadian Women's HIV Study Group. Viral polymorphism in human papillomavirus types 33 and 35 and persistent and transient infection in the genital tract of women. J Infect Dis. 2004;190(9):1575–85. doi: 10.1086/424854.
41. Godi A, Martinelli M, Haque M, Li S, Zhao Q, Xia N, et al. Impact of Naturally Occurring Variation in the Human Papillomavirus 58 Capsid Proteins on Recognition by Type-Specific Neutralizing Antibodies. J Infect Dis. 2018;218(10):1611–1621. doi: 10.1093/infdis/jiy354.
42. Mariaggi AA, Péré H, Perrier M, Visseaux B, Collin G, Veyer D, et al. Presence of Human Papillomavirus (HPV) Apolipoprotein B Messenger RNA Editing, Catalytic Polypeptide-Like 3 (APOBEC)-Related Minority Variants in HPV-16 Genomes From Anal and Cervical Samples but Not in HPV-52 and HPV-58. J Infect Dis. 2018;218(7):1027–1036. doi: 10.1093/infdis/jiy287.
43. Nejo YT, Olaleye DO, Odaibo GN. Molecular characterisation of genital human papillomavirus among women in Southwestern, Nigeria. PLoS One. 2019;14(11):e0224748. doi:

10.1371/journal.pone.0224748.

Figures

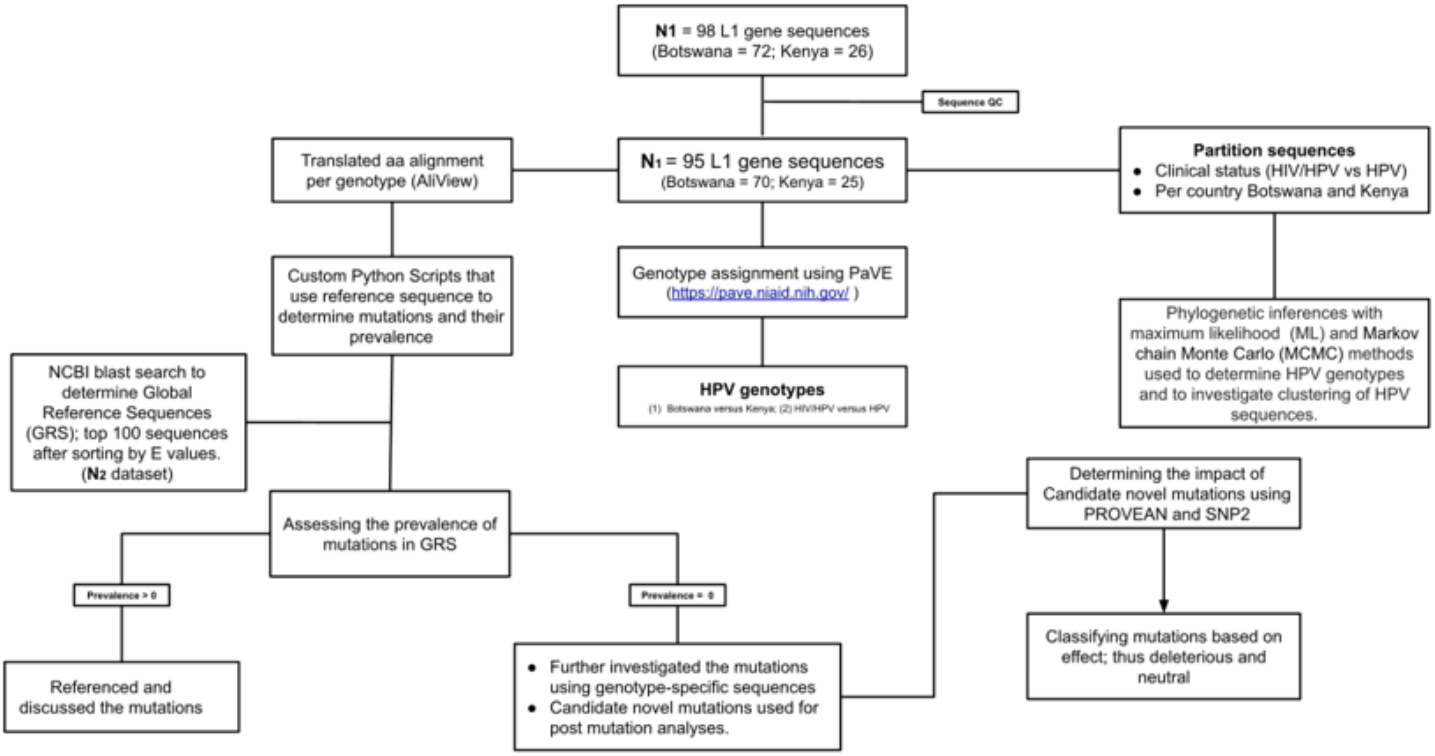

Figure 1

Scheme used to achieve study objectives

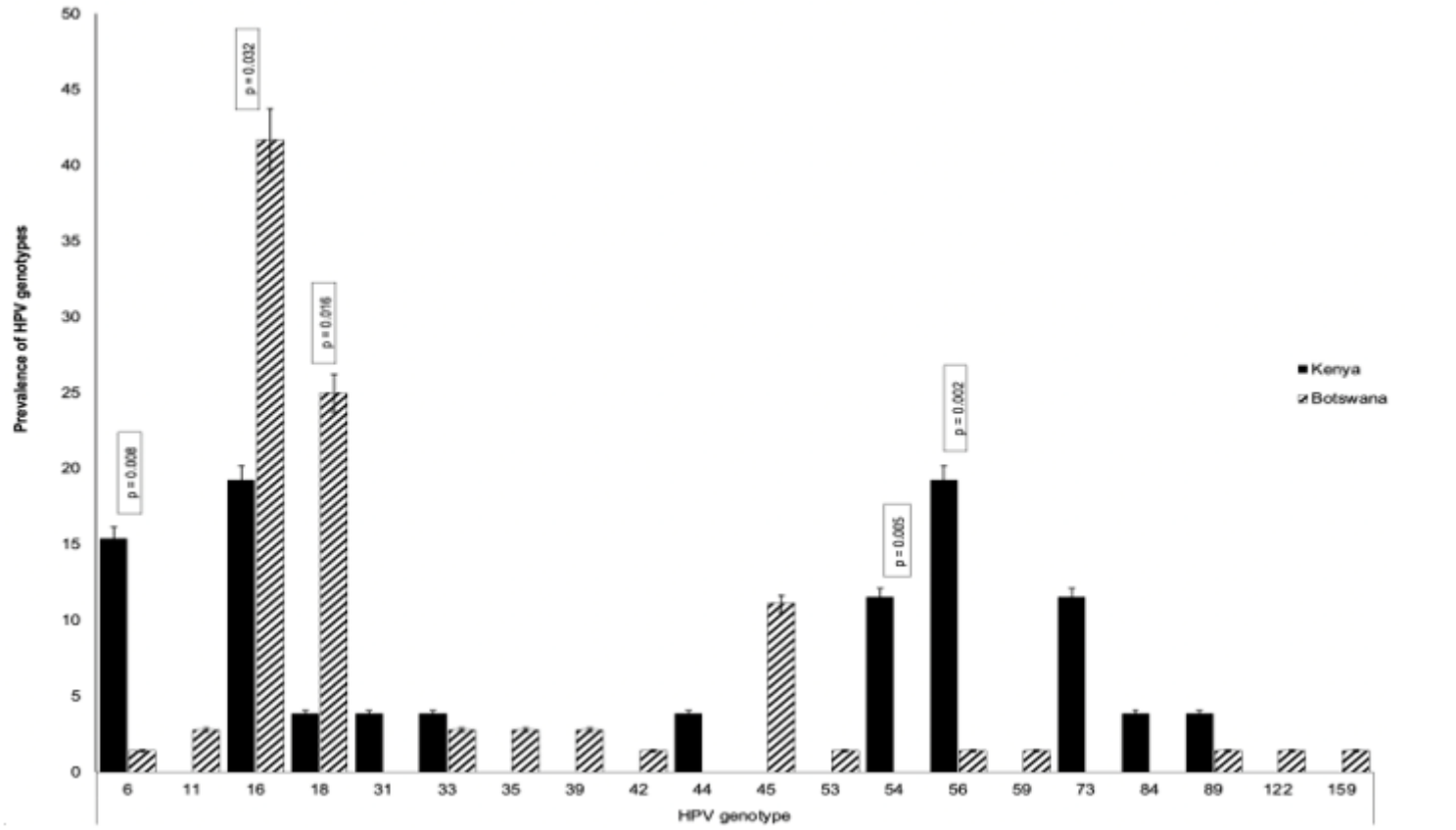

Figure 2

HPV genotypes in population of Botswana and Kenya

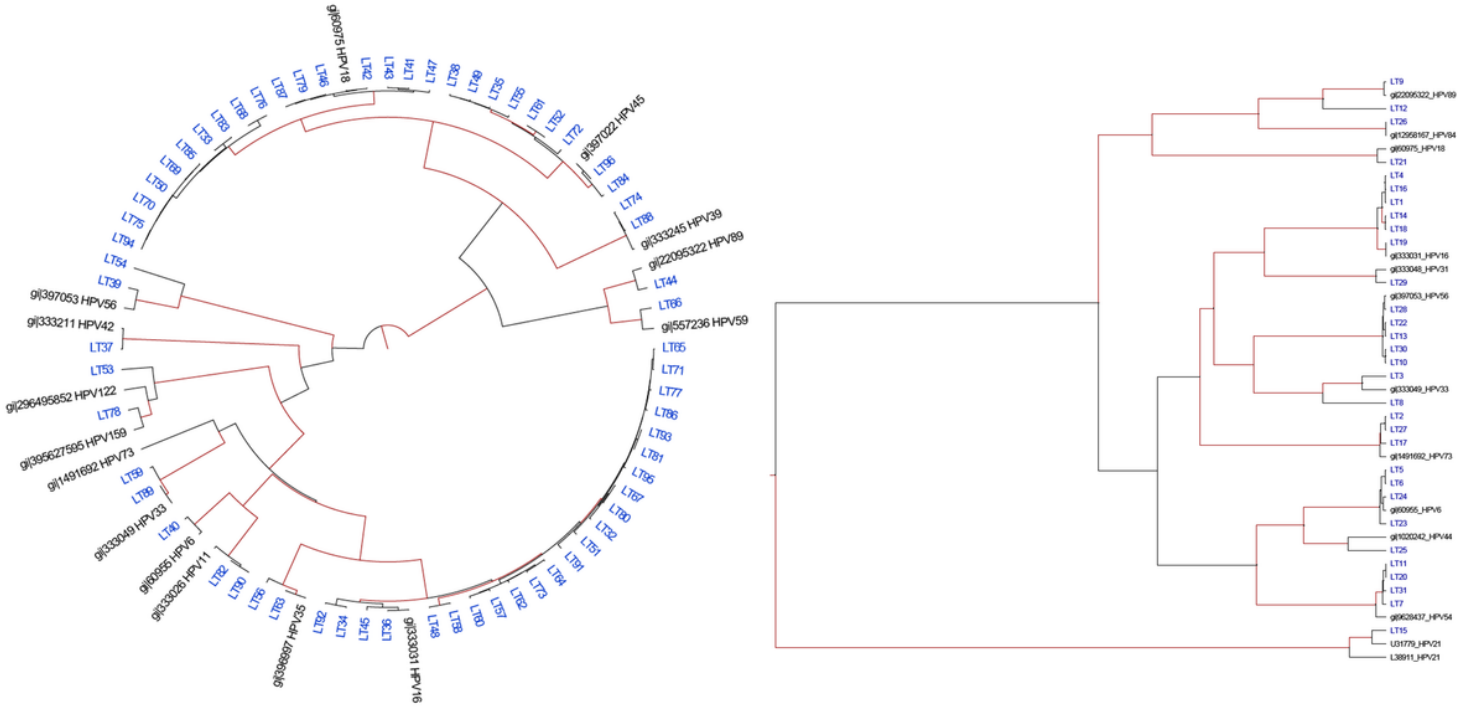

Figure 3

A phylogenetic tree of the L1 region (~450bp) covering nucleotide positions (5722- 6162) numbered to NC001526 HPV-16 reference genome) used to assign HPV types to sequences in this study. Trees were using BEAST method. Strains from Botswana sequenced in the present study are shown in the tree (left), while Kenya sequences are shown in the tree (right). Reference strains are designated by their accession number. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There were a total of 339 positions in the final dataset. Evolutionary analyses were conducted in BEAST v1.2.

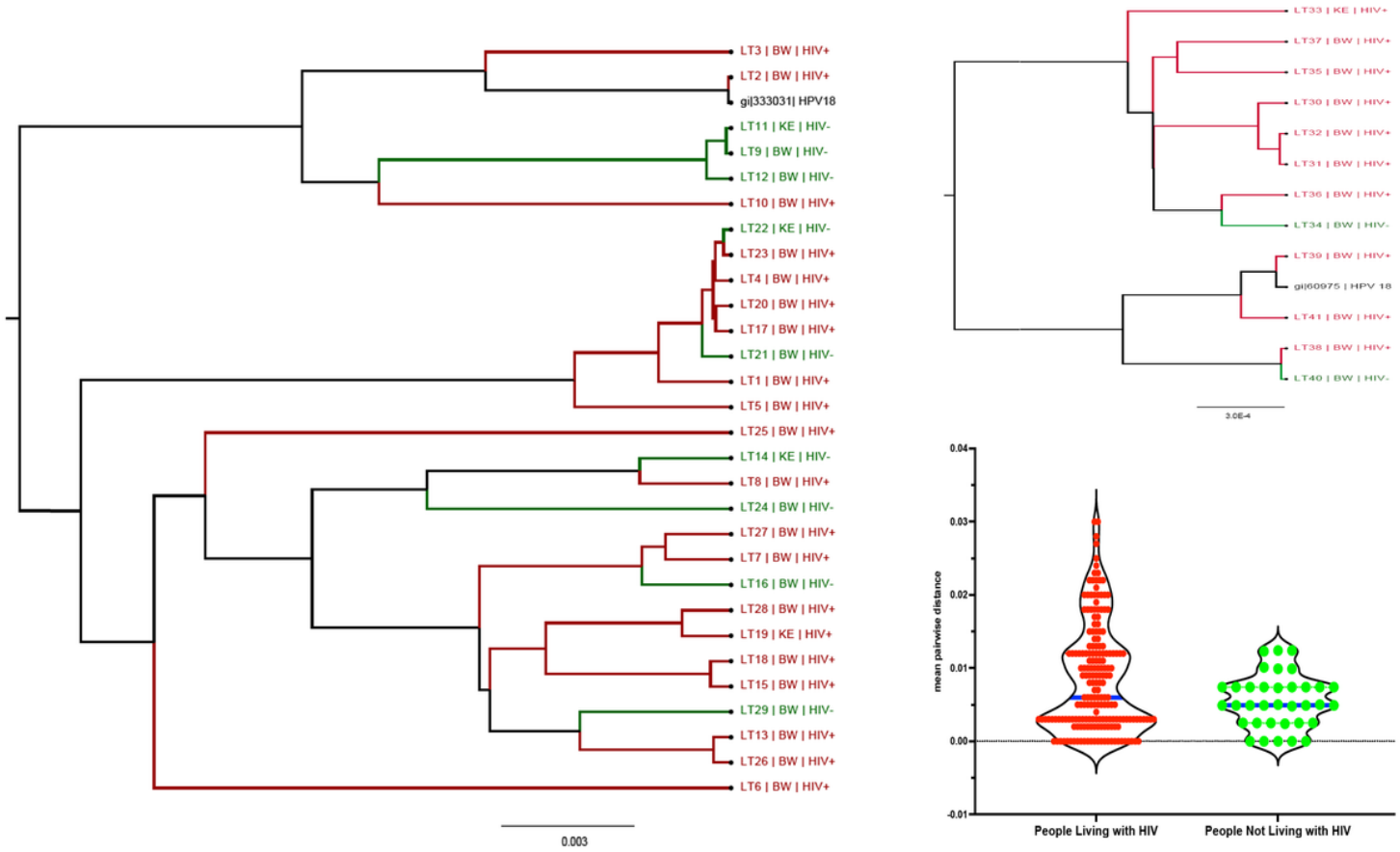

Figure 4

Phylogenetic analysis using BEAST tree for HPV-16 (A) and -18 (B) sequences from the present study. Mean per wise distribution of HPV 16 sequences isolated from WLWHIV versus WNLWHIV (C).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Tawesupplementarymaterial.doc