# Genetic analysis of human papillomavirus type 11 

 isolates from patients with recurrent respiratory papillomatosis treated at Universitas Academic HospitalCorne Thuynsma

July 2021


# Genetic analysis of human papillomavirus type 11 isolates from patients with recurrent respiratory papillomatosis treated at Universitas Academic Hospital 

by

## Corne Thuynsma

2014035048

Submitted in fulfilment of the requirements for the degree

Magister Medical Scientiae (Medical Virology)
in the Division of Virology
in the Faculty of Health Sciences
at the University of the Free State

# Supervisor: Prof. Felicity Jane Burt, Division of Virology, Faculty of Health Sciences, University of the Free State <br> Co-Supervisor: Prof. Riaz Seedat, Department of Otorhinolaryngology, Faculty of the Health Sciences, University of the Free State 

2021
Bloemfontein
South Africa

## Declaration

"I, Corne Thuynsma, declare that the Master's Degree research dissertation or interrelated, publishable manuscripts/published articles, or coursework Master's Degree mini-dissertation that I herewith submit for the Master's Degree qualification in Medical Virology at the University of the Free State is my independent work, and that I have not previously submitted it for a qualification at another institution of higher education."


Date: $\qquad$

## Acknowledgements

First and foremost, I would like to give my special regards to Prof. Felicity Burt. She offered invaluable advice and constant support during my M.Med.Sc study. Her immense knowledge and experience have encouraged me in my academic research and daily life. I wish to show my gratitude to Prof Riaz Seedat for his insightful comments and suggestions on my dissertation. I would also like to thank Armand for his support on the technical part of my study. You all provided invaluable assistance during my research.

My gratitude extends to the Poliomyelitis Research Foundation for the funding opportunity to undertake my studies at the Division of Virology, University of the Free State.

I wish to acknowledge the support and great love of my family and friends. To my laboratory matesGerdus, Micah, and Nina, I treasure the time spent together in the laboratory and the office. Coffee breaks with them have made my study and life in the Bloemfontein wonderful.

I want to express my gratitude to my family and loved ones for giving me all the wonderful opportunities that have made me who I am and helped me flourish. I am so lucky to have a family that fills my heart with strength, joy, and love. Without their tremendous encouragement, it would have been impossible for me to complete my studies.

Scientific research is one of the most exciting and rewarding of occupations - Frederick Sanger

## Table of Contents

List of figures ..... viii
List of tables ..... ix
List of abbreviations ..... xi
Abstract ..... $x v$
CHAPTER 1 - Literature review ..... 1
1.1. Background ..... 1
1.2. Discovery of human papillomavirus ..... 2
1.3. Human papillomavirus genome and proteins ..... 4
1.3.1. The early region ..... 4
1.3.2. The late region ..... 8
1.3.3. Upper regulatory region ..... 8
1.4. Classification and taxonomy of human papillomavirus ..... 9
1.5. Life cycle of human papillomavirus ..... 10
1.6. Transmission of human papillomavirus ..... 12
1.7. Epidemiology of human papillomavirus type 11 in South Africa ..... 13
1.8. Human papillomavirus pathogenicity ..... 14
1.9. Host immune response to human papillomavirus ..... 15
1.10. Recurrent respiratory papillomatosis ..... 16
1.10.1. Risk factors associated with the development of recurrent respiratory papillomatosis ..... 17
1.10.2. Diagnosis of recurrent respiratory papillomatosis ..... 18
1.10.3. Treatment of recurrent respiratory papillomatosis ..... 18
1.11. Prophylactic vaccines against human papillomavirus ..... 19
1.12. Detection of human papillomavirus infections ..... 21
1.12.1. Nucleic acid hybridisation assays ..... 21
1.12.2. Signal amplification assays ..... 21
1.12.3. Nucleic acid amplification assays ..... 21
1.12.4. Genotyping ..... 22
1.13. Phylogenetic analysis of human papillomavirus isolates ..... 23
1.13.1. Sequence evolution ..... 23
1.13.2. Multiple sequence alignment ..... 24
1.13.3. Phylogenetic trees ..... 24
1.13.4. Types of phylogenetic methods ..... 24
1.14. Problem statement ..... 25
1.15. Aims and objectives ..... 25
CHAPTER 2- Phylogenetic analysis of human papillomavirus type 11 isolates ..... 27
2.1. Introduction ..... 27
2.2. Materials and methods ..... 29
2.2.1 Study samples ..... 29
2.2.2 Study setting and population ..... 29
2.2.3 Amplification of five human papillomavirus genomic regions ..... 29
2.2.4 Agarose gel electrophoresis ..... 31
2.2.5 Purification of PCR product ..... 32
2.2.6 Determination of DNA concentration ..... 33
2.2.7 Sequencing ..... 33
2.2.8 Data analysis ..... 35
Analysis of nucleotide variation ..... 35
Degree of divergence ..... 35
Analysis of amino acid variation ..... 36
Phylogenetic analysis ..... 36
2.2.9 Ethical considerations ..... 36
2.3. Results ..... 37
2.3.1 Patient data ..... 37
2.3.2 L1 ORF amplification and sequencing ..... 37
2.3.3 URR amplification and sequencing ..... 40
2.3.1 E5a/b ORF amplification and sequencing ..... 42
2.3.2 E2 ORF amplification and sequencing ..... 44
2.3.3 Analysis of amino acid variations ..... 44
Analysis of L1 ORF sequences ..... 44
Analysis of E5a/b ORF sequences ..... 46
2.3.4 Phylogenetic analysis ..... 46
Maximum Likelihood method ..... 47
Pairwise analysis of percentage divergence of nucleotides ..... 49
Analysis of E2 segment as representative of whole genome variation ..... 51
2.4. Discussion ..... 55
CHAPTER 3- Characterisation of novel human papillomavirus type 11 isolates ..... 61
3.1. Introduction ..... 61
3.2. Materials and methods ..... 62
3.2.1. Study samples ..... 62
3.2.2. Full-genome amplification and sequencing ..... 63
3.2.3. Agarose gel electrophoresis ..... 65
3.2.4. Purification of PCR product ..... 65
3.2.5. Determination of DNA concentration ..... 65
3.2.6. MiSeq library preparation and sequencing ..... 65
3.2.7. Next-generation sequencing data analysis ..... 66
Analysis of nucleotide and amino acid variation ..... 66
Percentage variation in coding regions ..... 66
Degree of divergence ..... 66
Maximum likelihood method ..... 67
3.3. Results ..... 67
3.3.1. Selection of isolates for whole genome sequencing ..... 67
3.3.2. Nucleotide and amino acid variances across the human papillomavirus type 11 genome ..... 69
3.3.3. Pairwise nucleotide difference between complete human papillomavirus type 11 genomes ..... 72
3.3.4. Maximum Likelihood method ..... 73
3.3.5. Genome heterogeneities with respect to the corresponding prototype ..... 73
3.4. Discussion ..... 76
Chapter 4-Concluding remarks ..... 78
References ..... 82
Appendices ..... I
Appendix 1: Human papillomavirus type 11 positive patient sample information ..... I
Appendix 2: Human papillomavirus type 11 isolates retrieved from GenBank and accession numbers ..... IV
Appendix 3: Health Sciences Research Ethics Committee approval document ..... VI
Appendix 4: Design of E2 segment primers ..... VII
Appendix 5: Specificity of E2 segment primers ..... VIII
Appendix 6: 1x TAE preparation ..... IX
Appendix 7: Agarose gel preparation ..... IX
Appendix 8: GelRed stain preparation ..... IX
Appendix 9: Amino acid codon chart ..... X
Appendix 10: Pairwise analysis of percentage divergence of nucleotides using the human papillomavirus type 11 concatenated E5a/b-L1-URR data set
Appendix 11: Pairwise analysis of percentage divergence of nucleotides using the 208bp human papillomavirus type 11 E 2 segment data set XV
Appendix 12: Pairwise analysis of percentage divergence of nucleotides using the human papillomavirus type 11 complete genomes

## List of figures

Figure 1: Genome organisation of a low-risk human papillomavirus type 11 (HPV11).

Figure 2: Schematic drawing of human papillomavirus type 11 (HPV11) replication initiation mechanisms.

Figure 3: Human papillomavirus (HPV) replication in epithelial cells.

Figure 4: Phylogenetic tree of human papillomavirus type 11 (HPV11) based on alignment of 96 nucleotide sequences of HPV11 concatenated L1-URR. This analysis involves 68 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.

Figure 5: Phylogenetic tree of Human papillomavirus type 11 (HPV11) based on alignment of 95 nucleotide sequences of HPV11 concatenated E5a/b-L1-URR. This analysis involves 67 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.

Figure 6: Phylogenetic tree of Human papillomavirus type 11 (HPV11) based on 95 nucleotide sequence alignments of HPV11 E2 segment. This analysis involves 67 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.

Figure 7: Phylogenetic tree of Human papillomavirus type 11 (HPV11) based on 95 nucleotide sequence alignments of HPV11 E2 segment. This analysis involves 67 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.

Figure 8: Genome organisation of a low-risk human papillomavirus type 11 (HPV11).

Figure 9: Phylogenetic tree of human papillomavirus type 11 (HPV11) based 32 nucleotide sequence alignments of HPV11 including four isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.

## List of tables

Table 1: List of primers used for sequencing the L1, E5a/b, upper regulatory region (URR) and a segment of the E2 of human papillomavirus type 11 (HPV11) modified from Maver et al., 2011. 31

Table 2: Polymerase chain reaction (PCR) components for amplifying human papillomavirus type 11 (HPV11) L1 gene, E5a/b gene, and upper regulatory region (URR). 32

Table 3: List of primers used for sequencing the L1 gene of human papillomavirus type 11 (HPV11) from Maver et al., 2011.

Table 4: Sequencing reaction components for nucleotide determination of human papillomavirus type 11 (HPV11) isolates.

Table 5: Quantitative estimation of DNA concentration revealed by Nanodrop spectrophotometer of human papillomavirus type 11 (HPV11) genomic regions including late region L1, upper regulatory region (URR), early region E5a/b and a segment of early region E2.

Table 6: Sequence comparison of nucleotide variances among 72 human papillomavirus type 11 (HPV11) isolates L1 region against the HPV11 prototype with GenBank accession number M14119.1.

Table 7: Sequence comparison of nucleotide 68 among various human papillomavirus type 11 (HPV11) isolates upper regulatory region (URR) against the HPV11 prototype with GenBank accession number M14119.1.

Table 8: Sequence comparison of nucleotide variances among 67 human papillomavirus type 11 (HPV11) isolates E5a/b region against the HPV11 prototype with GenBank accession number M14119.1.43

Table 9: Sequence comparison of nucleotide variances among 67 human papillomavirus type 11 (HPV11) isolates E2 segment against the HPV11 prototype with GenBank accession number M14119.1.45

Table 10: GenBank accession numbers of human papillomavirus type 11 (HPV11) isolates and the (sub)lineage representation.

Table 11: List of primers used for sequencing full-length human papillomavirus type 11 (HPV11) genome in two overlapping fragments.

Table 12: Polymerase chain reaction (PCR) components for amplifying human papillomavirus type 11 full-length genome in two overlapping fragments (E1 to L1 genes; L1 to E1 genes). 64

Table 13: Patient information.

Table 14: Next-generation sequencing assembly report for four complete human papillomavirus type 11 (HPV11) mapped to the A1 prototype M14119.1 using Geneious Prime software (https://www.geneious.com).

Table 15: The lengths of the complete human papillomavirus type 11 (HPV11) consensus sequence, the CG content and the pairwise identity of four HPV11 isolates mapped to the A1 prototype M14119.1 using Geneious Prime software (https://www.geneious.com). 68

Table 16: Sequence comparison of nucleotide variances among four human papillomavirus type 11 (HPV11) isolates against the HPV11 prototype with GenBank accession number M14119.1. 70

Table 17: Comparison of amino acid variances among four human papillomavirus type 11 (HPV11) isolates against the HPV11 prototype with GenBank accession number M14119.1. 71

Table 18: Estimates of evolutionary divergence between four human papillomavirus type 11 (HPV11) sequences using MEGA X software.

Table 19: Genomic variations between human papillomavirus type 11 (HPV11) lineage B isolates VBD28/14 sequenced in this study and lineage B prototype LN833183.1 retrieved from GenBank. 73

Table 20: Genomic variations between human papillomavirus type 11 (HPV11) sub-lineage A2 isolates sequenced in this study and sub-lineage A2 prototype LN833161.1 retrieved from GenBank. 74

## List of abbreviations

| A | Alanine |
| :---: | :---: |
| AE | Early polyA signal |
| AIDS | Acquired immunodeficiency syndrome |
| ATP | Adenosine triphosphate |
| BLAST | Basic Local Alignment Search Tool |
| bp | Base pair(s) |
| CANSA | Cancer Association of South Africa |
| CD | Cluster of differentiation |
| CDK2 | Cyclin-dependent kinase 2 |
| D | Aspartic acid |
| DBD | Deoxyribonucleic acid binding domain |
| DC | Dendritic cells |
| ddNTPs | Dideoxynucleotides triphosphates |
| DNA | Deoxyribonucleic acid |
| ds | Double-stranded |
| E | Glutamic acid |
| E2F | E2 factor |
| E6AP | E6-associated protein |
| EDTA | Ethylenediaminetetraacetic acid |
| F | Phenylalanine |
| FTR | Formylmethanofuran-tetrahydromethanopterin formyltransferase |


| G | Glysine |
| :---: | :---: |
| G1-phase | Growth-1-phase |
| G2-phase | Growth-2-phase |
| H | Hinge region |
| hc2 | Hybrid Capture®2 |
| HD | Helicase domain |
| HIV | Human immunodeficiency virus |
| HPV(s) | Human papillomavirus(es) |
| HR | High-risk |
| HSPGs | Heparin sulphate proteoglycans |
| I | Isoleucine |
| ICTV | International Committee on Taxonomy of Viruses |
| IgG | Immunoglobulin-G |
| K | Lysine |
| L | Leucine |
| LCR | Long control region |
| LR | Low-risk |
| MCHA | Microplate colorimetric hybridisation assay |
| MEGA | Molecular Evolutionary Genetics Analysis |
| MHC | Major histocompatibility complex |
| M-phase | Mitosis-phase |
| MUM1+ | Multiple myeloma oncogene-1 |


| N | Asparagine |
| :---: | :---: |
| NCR | Non-coding region |
| NGS | Next generation sequencing |
| OBD | Origin binding domain |
| ORF(s) | Open reading frame(s) |
| $\mathbf{P}$ | Proline |
| p53 | Tumour protein 53 |
| PCR | Polymerase chain reaction |
| pRB | Retinoblastoma protein |
| Q | Glutamine |
| $\mathbf{R}$ | Arginine |
| rcf | Relative centrifugal force |
| RFLP | Restriction-fragment length polymorphism |
| RNA | Ribonucleic acid |
| RRP | Recurrent respiratory papillomatosis |
| S | Serine |
| SIN | Squamous intraepithelial neoplasia |
| SNP(s) | Single nucleotide polymorphism(s) |
| S-phase | Synthesis-phase |
| STI | Sexually transmitted infection |
| T | Threonine |
| TAD | Transactivation domain |


| TAE | Tris-Acetic acid- ethylenediaminetetraacetic acid |
| :--- | :--- |
| Tm | Melting temperature |
| UFS | University of the Free State |
| URR | Upstream regulatory region |
| V | Volts |
| VLP(s) | Virus-like particle(s) |
| WGS | Whole genome sequencing |


#### Abstract

Human papillomavirus type 11 (HPV11) is a causative agent of recurrent respiratory papillomatosis (RRP), a common benign laryngeal neoplasm that presents mainly in children. The genome comprises three regions: the early region (E1, E2, E4, E5a/b, E6 and E7), the late region (L1 and L2), and the upper regulatory region (URR). A sequence-based classification system is primarily used to genotype HPV. The L1 is used for HPV type discrimination, and in combination with the URR, can be used to differentiate between various lineages. However, optimal sub-lineage classification requires whole genome sequencing (WGS). A recent study investigating the genomic diversity of globally circulating HPV11 isolates identified a novel lineage and two novel sub-lineages. It has been proposed that phylogenetic tree topologies using the sequences of concatenated E5a/b- L1-URR genes, a 208bp segment of the E2 gene, and the complete genome generates similar tree topologies. Also, there is currently no published data on the HPV11 intratypic variants circulating in the Free State region. Hence, this study investigated HPV11 intratypic variants circulating in patients with RRP at the Universitas Academic Hospital, and aimed to identify novel (sub)lineages through phylogenetic investigations.

The study population included patients diagnosed with RRP caused by HPV11, and sequence data for geographically distinct HPV11 (sub)lineage representatives. The genetic variation of HPV11 isolated from patients with RRP was determined by sequencing the E5a/b, L1, URR and a segment of E2 genes. Four isolates of interest were selected for whole-genome sequencing and phylogenetically analysed to determine the presence of potentially novel isolates.

Many nucleic heterogeneities and non-synonymous substitutions were identified in isolates characterised in this study. Phylogenetic analysis of the concatenated L1-URR and E5a/b-L1-URR resolved into lineages A and B; however, sub-lineage classification was unclear. Analysis of the complete genome determined the presence of a lineage B isolate and two isolates of interest. Comparative analysis of genetic variability determined that the concatenated E5a/b-L1-URR could not reliably classify isolates. A segment of the E2 gene could reliably distinguish between all lineages and sub-lineages, suggesting that this gene segment contains stable sub-lineage specific single nucleotide polymorphisms (SNPs) and may serve in sub-lineage identification.

In conclusion this study provides the most comprehensive data on the genomic diversity of HPV11 in the Free State to date. Results obtained in the current study support WGS for HPV11 classification below lineage level as a standard, as it generates more information regarding genetic variants.


## CHAPTER 1 - Literature review

### 1.1. Background

The first papillomavirus was identified in the 1930s (Shope \& Hurst, 1933). Since then, recombinant deoxyribonucleic acid (DNA) technology and molecular cloning have led to the discovery of multiple human papillomaviruses (HPVs) (Boshart et al., 1984; Dürst et al., 1983; Gissmann et al., 1982). Papillomaviruses target various host tissues and are responsible for forming benign hyperproliferation and malignancies in human and animal hosts. The association with different conditions such as cancer, skin warts and laryngeal papillomas has been well-studied (Cornall et al., 2013; Gaylis \& Hayden., 1991; zur Hausen et al., 1975).

HPVs which target mucosal sites are broadly divided into high-risk (HR)-HPV types, probable HRHPV types and low-risk (LR)-HPV types based on oncogenic potential (zur Hausen, 2002). HPV6 and HPV11 are LR-HPV types responsible for genital wart formation and development of papillomas in the respiratory tract, commonly referred to as recurrent respiratory papillomatosis (RRP) (Kardani \& Bolhassani, 2018b). This disease is characterised by recurrent papillomas forming in the respiratory tract, most commonly the larynx and may cause significant morbidity and possible mortality due to airway obstruction, especially in younger children (Seedat, 2020; Seedat et al., 2013; Swain et al., 2020).

HPVs form part of the Papillomaviridae family and are non-enveloped, double-stranded (ds) deoxynucleic acid (DNA) viruses containing a small genome of around 8000 base pairs (bp) (Crawford, 1965; Klug \& Finch, 1965). The virus genome comprises three regions, which include the early region (E1-E7 genes) and the late region (L1 and L2 genes) that encode proteins, and the upstream regulatory region (URR), also recognised as the long control region (LCR) (Kardani \& Bolhassani, 2018a).

A classification and nomenclature system for the Alphapapillomaviruses was established in 2011 by Burk and colleagues. Papillomavirus classification largely relies on phylogenetic analysis, genome organisation properties, virulence factors and the host tissue the virus affects. A DNA sequence-based classification system is primarily used to differentiate and genotype HPV (Burk et al., 2011). The L1 open reading frame (ORF) is generally used for HPV type discrimination, and in combination with the URR, can be used to differentiate between various lineages. However, optimal sub-lineage classification requires sequencing additional gene segments (Burk et al., 2011). A novel HPV type identifies as having over $10 \%$ variance to any other HPV type. Nucleotide differences in the complete genome of $1 \%$ to $10 \%$ and $0.5 \%$ to $1 \%$ identify variant lineages and sub-lineages, respectively (Burk et al., 2011).

DNA sequencing methods, such as Sanger sequencing and next generation sequencing (NGS), are often used in phylogenetic studies to investigate HPVs inheritable traits within different organisms and to detect genomic mutations. By analysing the data acquired by sequencing, phylogenetic trees, which reflect the evolutionary history of organisms, can be constructed, and the information gathered can then be used for taxonomic classification and analyses (Burk et al., 2011; Kocjan et al., 2015; Sridhar et al., 2015). Using phylogenetic analysis, HPV11 was previously divided into subgroups A1 and A2; however, recently a study on the genomic diversity of globally circulating HPV11 isolates uncovered two lineages (A and B) and four sub-lineages (A1, A2, A3 and A4) (Jelen et al., 2016), warranting investigation of further isolates for clarification of lineages currently circulating (Burk et al., 2011).

Previous studies conducted on young adults (age $16-24$ years) in South Africa determined that over two-thirds of the population are often infected with multiple HPV types, with approximately 1-5\% of the HPV positive participants infected with HPV11 (Giuliano et al., 2015; Mbulawa et al., 2017, 2018). A study conducted in the Free State province of South Africa determined that RRP has a predominantly juvenile-onset (children age 14 and under) ( $\sim 85 \%$ ), with approximately $40 \%$ of cases being due to HPV11 (Seedat \& Schall, 2018). However, most of these studies only target selected population groups and an accurate HPV11 prevalence, especially in the respiratory tract, is unknown and presumed to be much higher.

To date, there is no cure for RRP, and patients often require multiple surgical procedures to remove the papillomas to prevent airway obstruction (Seedat, 2020). Fortunately, preventative measures are available. There are three vaccines used for protection against HPV: Cervarix®, Gardasil® and Gardasil-9®. Two of these vaccines, Gardasil® and Gardasil-9®, protect against HPV11 infection when administered prophylactically (Barra et al., 2019; Chabeda et al., 2018; Dadar et al., 2018). To combat infection with HPV, the National Department of Health has incorporated HPV vaccination with Cervarix® into the school health programs, targeting half a million girls (Delany-Moretlwe et al., 2018). However, Cervarix ${ }^{\circledR}$ does not confer protection against HPV11 infection (Barra et al., 2019; Chabeda et al., 2018; Dadar et al., 2018).

Research on the HPV11 lineages and sub-lineages circulating within the community is urgently required to further study lineage correlation with disease severity, to guide vaccine development, to monitor the impact of the vaccination campaign on the circulating HPV types, to collect baseline data for future studies, and to identify novel HPV types if present.

### 1.2. Discovery of human papillomavirus

In the early 1930s, Rous and Shope identified the first papillomavirus from wart-like lesions in cottontail rabbits (Sylvilagus sp.), and further investigations led to the discovery of papillomavirus
malignant- and transmission capabilities (Rous \& Beard, 1935; Shope \& Hurst, 1933). Following this discovery, research into papillomaviruses intensified. In 1949, the first virus particles in a human papilloma extract were identified using electron microscopy (Strauss et al., 1949), and 16 years later, in 1965, the human wart virus's genome was determined to be circular, dsDNA with a protein capsid comprised of 72 pentameric capsomers (Crawford, 1965; Klug \& Finch, 1965).

Until the 1970s, it was believed that there was only a single type of HPV that caused the formation of warts in various tissue sites. However, within the decade, recombinant DNA technology and molecular cloning techniques led to the discovery of a plurality of HPV types, such as HPV6, HPV11, HPV16 and HPV18, with different tropisms for mucosal or cutaneous squamous surfaces (Boshart et al., 1984; Dürst et al., 1983; Gissmann et al., 1982). In the wake of these studies, the involvement of HPV in malignant cell transformation was hypothesised and would later prove to be a major focus area in HPV research (Gissmann et al., 1977; Gissmann \& zur Hausen, 1976; zur Hausen et al., 1975).

In 1982, molecular cloning of papillomavirus genomic DNA extracted from a respiratory papilloma led to the discovery of HPV11 (Gissmann et al., 1982). Further investigations eventually revealed that HPV11 was also responsible for genital wart development (Gissmann et al., 1983). The first complete nucleotide sequences of ten HPV11 isolates were determined and characterised in 1986 in Slovenia, and approximately a decade later, the URR sequences of 40 HPV11 isolates were determined, providing further insight into the genetic code of HPV11 types and intratypic variants (Dartmann et al., 1986; Heinzel et al., 1995). Only again in 2009, was the complete sequence of HPV11 obtained from a cervical swab of a patient with an unspecified genital disease (Wu et al., 2009).

The first extensive sequencing study to obtain complete genome sequences of HPV11 isolates was performed in 2011 in Slovenia (Maver et al., 2011). Phylogenetic analysis revealed sequences corresponding to the prototypic- and non-prototypic variant group based on specific nucleotide signatures and several novel HPV types and potentially significant mutations (Maver et al., 2011). The prototypic variant group and the non-prototypic variant group were renamed A1 and A2, respectively, when a new system for the classification and nomenclature was described for HPV variants (Burk et al., 2011). More recently, a study investigating the genomic diversity of globally circulating HPV11 isolates revealed two lineages (A and B) and four sub-lineages (A1 - A4) (Jelen et al., 2016). It has been established that HPV has a slow evolutionary rate, and genomic variation is predominantly a consequence of genetic drift (Bernard, 1994; van Doorslaer, 2013). Therefore, discovering any new HPV lineages or sub-lineages requires additional investigation into the evolutionary rate and the genetic relationship of HPV11 variants to further characterise these variants.

### 1.3. Human papillomavirus genome and proteins

HPVs are non-enveloped, dsDNA viruses with a small genome of around 8000bp (Kardani \& Bolhassani, 2018a). Many proteins necessary for viral infection, cellular gene expression, and immune evasion are encoded by the HPV ORFs. The genome comprises three regions, including the early ORF, late ORF, and the URR (Figure 1). The early region is approximately 4000bp, and codes for nonstructural proteins, whereas the late region is approximately 3000 bp , and codes for structural proteins (Kardani \& Bolhassani, 2018a). The URR is approximately 1000 bp and contains elements involved in regulating viral replication and gene expression (Harari et al., 2014; Kardani \& Bolhassani, 2018a; Ribeiro et al., 2018).

Throughout evolution, papillomaviruses have gained and lost protein coding genes; however, all papillomaviruses encode for a minimum of five proteins, including E1, E2, E4, L1 and L2, and contain a URR (Kardani \& Bolhassani, 2018a). Therefore, a virus containing only the basic set of core genes (E1, E2, L1, L2 and URR) should theoretically be able to infect a cell and replicate. It is thus hypothesised that ancestral papillomaviruses comprise the essential genes and no additional adaptive proteins. All papillomaviruses contain at least one adaptive protein, such as E5, E6 or E7, which plays a role in immune evasion and cellular growth (Kardani \& Bolhassani, 2018a, 2018b). Moreover, genes that encode adaptive proteins such as E5, E6 and E7, have been associated with malignant transformation, although not all papillomaviruses encode for all these adaptive proteins, and not all papillomaviruses have an association with malignant transformation (Egawa et al., 2015; van Doorslaer, 2013).

### 1.3.1. The early region

The early region comprises several ORFs that encode non-structural proteins, which play a regulatory function immediately following infection. These ORFs can be divided into the regulatory genes (E1, E2 and E4), which play a crucial role in virion synthesis, and three oncoproteins (E5, E6 and E7). The ORFs in the early region are expressed within the deeper, less differentiated layers of the infected tissue, such as the basal layer (Harari et al., 2014; Kardani \& Bolhassani, 2018a).

The E1 and E2 genes, play pivotal roles during replication and amplification of the viral circular dsDNA genome and are essential for virus survival. E1 is separated into three functional regions: The N terminal regulatory region, which stimulates cyclin-dependent kinase 2 (CDK2) phosphorylation, the central origin-binding domain (OBD), which binds E1 and E2 protein forming a complex (E1-E2), and the C-terminal helicase domain (HD) which acts as adenosine triphosphate (ATP)-dependent helicase (Graham \& Faizo, 2017; Kardani \& Bolhassani, 2018a). Functions of this protein include optimising viral reproduction in vivo, recognising the origin of replication, and assembling into a double hexameric
complex that unwinds DNA duplexes upstream of the DNA replication fork. E2 contains a DNAbinding domain and a protein-binding domain (Graham \& Faizo, 2017; Kardani \& Bolhassani, 2018a; Wallace \& Galloway, 2014).


Figure 1: Genome organisation of a low-risk human papillomavirus type 11 (HPV11).
E1-E7 early genes, L1-L2 late genes, URR Upper regulatory region, AE early polyA signal, and P97 and P742 promotors are indicated. The figure is drawn based on the HPV11 prototype A1 with a 2bp insertion at genomic position 7717-7718 (GenBank accession number M14119.1) (Maver et al., 2011).

The E2 protein is a multifunctional protein encoded by all papillomaviruses expressed during early and intermediate viral life cycle stages. The E2 has several primary functions essential for the viral life cycle, such as regulating expression levels for viral gene products in the early stages of the HPV life cycle, recruiting the E1 helicase protein to the viral origin of replication for replication initiation, and delivering the replicated viral genome to daughter cells during the division of the parent cell. Additionally, the E2 plays a role in transcription as a transcription factor and ensures low viral concentrations in the early stages of the HPV life cycle. Mutation or integration of the viral genome can result in the inactivation of the E2 protein, leading to over-expression of E6 and E7 genes (Graham \& Faizo, 2017; Kardani \& Bolhassani, 2018a; Wallace \& Galloway, 2014).

Replication is initiated when E2 recruits E1 to the viral origin (Figure 2). The recruitment step involves a crucial protein-protein interaction between the transactivation domain (TAD) of the E2 protein and the HD of E1. Then, additional E1 molecules are recruited, and the E2 protein promotes assembly into a replication-competent double hexameric helicase. ATP likewise promotes the oligomerisation of E1 and is further required to power E1 helicase activity. Lastly, E1 interacts with host cell replication factors to promote bidirectional replication of the viral genome (Graham \& Faizo, 2017; Kardani \& Bolhassani, 2018a; Wallace \& Galloway, 2014) (Figure 2).

The HPV E4 ORF is located within the E2 ORF and is the first gene expressed in the later stage of infection. E4 does not possess an initiation codon and uses the E2 initiation codon. The E4 protein plays a role in virion release by restructuring the epithelial cell's cytokeratin and may play a part in genome amplification and the enhancement of virion synthesis. The E4 protein displays various effects on cell behaviour, such as suppressing the host cell's DNA synthesis to promote apoptosis in terminally differentiated cells by interacting with the host mitochondria (Graham \& Faizo, 2017; Kardani \& Bolhassani, 2018a). In addition to this, it has also been reported to have a potential role in HPV screening (Yajid et al., 2017).

The E5 protein is a transmembrane hydrophobic protein expressed late in infection made up of approximately 83 amino acids. The protein has several functions, including immune evasion via major histocompatibility complex (MHC) repression and regulating apoptosis. Moreover, the E5 protein plays a role in cell cycle pathways via interaction with growth factor receptors (Graham \& Faizo, 2017; Kardani \& Bolhassani, 2018a). Similar to other HR proteins, the E5 influences cellular gene expression by regulating small non-coding ribonucleic acid (RNA) molecules, thereby exhibiting oncogenic activity (Georgescu et al., 2018).

The E6 and E7 proteins exhibit oncogenic activity by associating with tumour protein 53 (p53) and retinoblastoma protein (pRB) (Altamura et al., 2018; Georgescu et al., 2018; Pal \& Kundu, 2020). Expression of these ORFs in the lower epithelial layer triggers differentiating cells of the suprabasal layers (the thin layer above the basal layer) to re-enter the synthesis-phase (S-phase) in which DNA replicates. This phase is highly regulated and conserved, and dysregulation generally leads to cell apoptosis (Georgescu et al., 2018; Kardani \& Bolhassani, 2018a). E6 proteins are responsible for degrading p53, which inhibits apoptosis, and therefore the infected cells continue viral DNA replication by recruiting E1/E2 complexes which play a significant role in replication. The pRB gene acts as a tumour suppressor by suppressing the replication of enzyme expression genes (Pal \& Kundu, 2020). When E7 binds to pRB , the complex inactivates, stimulating replication and cell division. Association of the oncoproteins with these tumour suppressor genes leads to disruption of the physiological functions of these genes and subsequent possible malignant transformation (Pal \& Kundu, 2020).


Figure 2: Schematic drawing of human papillomavirus type 11 (HPV11) replication initiation mechanisms.
Early proteins E1 and E2 are involved. Locations of the functional domains in the proteins are indicated. OBD origin binding domain, TAD transactivation domain, DBD Deoxyribonucleic acid binding domain, and $H$ hinge region. Replication is initiated by E1 (red), by E2 (blue), to the OBD. This involves an interaction between the E2 TAD and the E1 helicase domain. E2 recruits additional E1 molecules. Adenosine triphosphate (ATP) also stimulates helicase activity of E1, which interacts with host cell replication factors to promote bidirectional replication of the viral genome. E1 interacts with host cell replication factors to promote bidirectional replication of the viral genome. Adapted from D'Abramo \& Archambault (2011) and Stenlund (2003).

Recent studies have demonstrated the role of oxidative stress and chronic inflammation in carcinogenic development. The oncogenes E5, E6 and E7, are all involved in developing chronic inflammation, which induces oxidative stress, leading to cell damage and subsequent malignant transformation (Georgescu et al., 2018). The type of papillomavirus influences the oncogenicity of E5, E6 and E7. HPV types such as HPV11 have limited activity compared to the other HPV types and are classified as LR-HPV. LR-HPV types, unlike HR-HPV types, do not degrade pRB or p53 and are less likely to result in malignant transformation than their HR counterparts (Egawa \& Doorbar, 2017; Klingelhutz \& Roman, 2012; Pal \& Kundu, 2020).

### 1.3.2. The late region

The late region contains two ORFs, namely L1 and L2, which express structural- or capsid proteins required for transmission and survival of HPV in the upper, more differentiated layers of the infected tissues (Buck et al., 2013; Wang \& Roden, 2013). The L1 major capsid protein can spontaneously selfassemble into virus-like particles (VLPs), presenting an icosahedral exterior surface. The L1 gene is highly conserved, but contains a small number of lineage specific nucleotide polymorphisms that makes it useful for typing (Buck et al., 2013; Burk et al., 2011). The structure is highly immunogenic and has formed the basis of successful vaccines targeted against HR-HPV types (Buck et al., 2013).

The L1 capsid surface contains knob-like structures constituting 360 copies of the protein organised into 72 pentameric capsomers with a copy of L2 protein in the centre. The L1 has two termini, namely the N - and C termini, arranged as 'invading arms' arranged to form the floor between the knob-like capsomeres (Buck et al., 2013; Conway \& Meyers, 2009). The capsid undergoes various essential conformational changes during the virus life cycle. These conformational changes mediate vital functions for virus survival, including encapsidation of the papillomavirus genome, interaction with the host cell for infectious entry, and releasing the viral DNA into a new host cell (Buck et al., 2013).

The capsid protein L2 has complex roles in the biology of all papillomaviruses, most notably in virus assembly and the infectious process. Unlike the major capsid protein L1, the minor capsid protein L2 cannot self-assemble into VLPs, although it can be integrated into VLPs when co-assembled and expressed with the L1 protein (Conway \& Meyers, 2009; Wang \& Roden, 2013). Roles of the L2 protein include facilitation of efficient genome encapsidation, vesicular trafficking of the viral genome in the direction of the host cell nucleus and escaping the vesicular compartment to travel to the host cell nucleus for successful infection (Conway \& Meyers, 2009; Wang \& Roden, 2013).

### 1.3.3. Upper regulatory region

The URR is the most variable region within the HPV genome since it does not encode for proteins. Consequently, it is capable of accumulating and tolerating more nucleic mutations. The URR is located between the E6 ORF and L1 ORF and consists of approximately 850bp (Kardani \& Bolhassani, 2018a; Ribeiro et al., 2018). The URR interacts with numerous cellular and viral factors and is involved with functions such as virus replication, gene expression, and transcription. The region divides into three segments which all include E2 binding sites. The URR includes the 5 ' segment, the central segment and the 3 ' segment. The 3 ' segment also bears the E1 binding site, which overlaps with the origin of replication (Graham \& Faizo, 2017; Ribeiro et al., 2018). The URR includes genetic elements, for example, early promotors, the transcription enhancer, the origin of replication, the late polyadenylation
site, and the late regulatory element. Mutations in the URR may impact binding sites and functionality (Fang et al., 2020; Graham \& Faizo, 2017; Ribeiro et al., 2018).

### 1.4. Classification and taxonomy of human papillomavirus

Initially, papillomaviruses and polyomaviruses were grouped into one family, namely the Papovaviridae. These viruses were subsequently separated into two families, Papillomaviridae and Polyomaviridae, due to the lack of nucleotide and amino acid similarities, different genome sizes and genome organisation (Bernard et al., 2010; Fauquet et al., 2005; van Regenmortel et al., 2002). Papillomaviruses consists of a family of viruses that classifies into different genera, species, and types. HPVs consist of three significant genera, including the Alphapapillomaviruses, Betapapillomaviruses and Gammapapillomaviruses, but smaller genera Mupapillomavirus and Nupapillomavirus, also exist (Bernard et al., 2010; de Villiers, 2013; Murahwa et al., 2019).

Throughout the evolutionary process, nucleotide sequences across the HPV genome have transformed. Evolutionary changes in papillomaviruses' hosts created new ecological niches for papillomaviruses to adapt to, resulting in the formation of different clades. These viruses then co-evolved alongside their specific hosts resulting in further co-speciation (Bernard, 1994; van Doorslaer, 2013). HPV tropism is used to classify viruses into cutaneous viruses or mucosal viruses. Mucosal viruses are further divided into LR, probable HR, and HR viral types (Egawa \& Doorbar, 2017; Egawa et al., 2015).

A DNA sequence-based classification system is primarily used to classify HPVs. The International Committee on Taxonomy of Viruses (ICTV) established the designation and naming of virus taxa based on recommendations from the Papillomaviridae Study Group, which consists of 11 researchers (https://talk.ictvonline.org/taxonomy/). A classification and nomenclature system for the Alphapapillomaviruses was established in 2011 by Burk and colleagues (2011). This study was the first to describe the nomenclature system for HPV6 and HPV11 variants based on whole-genome analyses.

The ORF nucleotide sequence coding for the capsid protein L1 is used to classify HPV types and represent whole-genome variation due to the high level of conservation of the L1 gene (Bernard et al., 2010; Burk et al., 2011; de Villiers et al., 2004; de Villiers, 2013). Consensus primers that target the L1 ORF are used to differentiate between various HPV types. These amplicons can be sequenced and aligned to existing HPV prototypes to identify HPV types, lineages and sub-lineages or any nucleic heterogeneity (Bernard et al., 2010; Burk et al., 2011; de Villiers et al., 2004; de Villiers, 2013). A novel HPV type has less than $90 \%$ similarity to any HPV type. Nucleotide differences in the complete genome of between $1.0 \%$ to $10.0 \%$ and $0.5 \%$ to $1.0 \%$ define variant lineages and sub-lineages, respectively (Burk et al., 2011). Data acquired by sequencing is used to construct phylogenetic trees and used for taxonomy (Burk et al., 2011). Approximately 220 HPV types have thus far been identified
as of 2021 according to the International HPV Reference Centre (Bernard et al., 2010; de Villiers et al., 2004) (https://www.hpvcenter.se/human_reference_clones).

The HPV11 genome was formally classified into two sub-lineages (A1 and A2) based on complete HPV11 genomes (Burk et al., 2011; Maver et al., 2011). A more recent study on the global genomic diversity of HPV11 identified a separate lineage (lineage B) and two additional sub-lineages (sublineages A3 and A4) (Jelen et al., 2016). These results were also generated by sequencing the E5a ORF, E5b ORF, L1 ORFs and the URR, which reliably produced the same phylogenetic results as wholegenome sequencing. Moreover, this study identified a 208bp region found between the 3' end of the E2 ORF and the 5 ' of the non-coding region- 2 (NCR2) that is the most suitable genomic region to sequence to distinguish all identified HPV11 lineages and sub-lineages when whole-genome sequencing is impractical (Jelen et al., 2016). A review of literature revealed that additional studies on HPV11 are required to investigate the heterogeneity observed in this study.

### 1.5. Life cycle of human papillomavirus

Papillomaviruses all follow similar life cycles but are species-specific regarding infection and replication within hosts (Doorbar et al., 2012; Kardani et al., 2018). They infect epithelial cells, which are part of the cutaneous and mucosal tissue that generally serves as a protection against the external environment (Madison, 2003; Stanley, 2020).

Basal cells compose the bottom layer of the epithelial layer and are capable of mitotic division. During mitosis, a daughter cell moves up through the epithelial layer to replace non-living surface epithelial cells. This movement marks the end of the cell cycle and the beginning of terminal differentiation, resulting in the loss of the nucleus (Kardani et al., 2018; McBride, 2017). The cell cycle occurs in four phases. Cellular growth occurs in the growth-1-phase (G1-phase), host DNA replicates in the S-phase, the cells prepare for division in the growth-2-phase (G2-phase), and chromosomes are separated, and daughter cells form in the mitosis-phase (M-phase) (Kardani et al., 2018; McBride, 2017). HPVs do not possess replication machinery apart from the E1 replication enzyme. To overcome this, HPVs depend on the epithelial cellular replicative enzymes for propagation and enter basal cells to link virus and host replication life cycles (Doorbar et al., 2012; McBride, 2017).

There are four main stages of replication during the HPV life cycle: virus entry, establishment, maintenance of the non-productive infectious state, and virus production (Kardani et al., 2018; McBride, 2017). HPV replication is depicted in Figure 3. The first phase of HPV replication is host cell entry. Infection with HPV is thought to require epithelium micro-abrasions to allow access of the virus into actively dividing basal cells; however, multiple entry-pathways have been reported for the immense diversity of HPV types (Graham \& Faizo, 2017; Kardani et al., 2018; McBride, 2017). The L1 capsid
protein then binds to the host cell receptor such as heparin sulphate proteoglycans (HSPGs) which are the primary receptors. Receptor types and strategies may be dependent on the HPV type and the host cell. The virus enters the host cell using endocytosis and travels through vesicular trafficking to the nucleus, where the virus enters the nucleus through nuclear pores or following cell mitosis (Graham \& Faizo, 2017; Kardani et al., 2018; McBride, 2017).

Once the virus has successfully entered the host cell, the second- or non-productive infectious state is established. The initial amplification phase of the virus is thought to be transient and rapid, and relies on host cell machinery for replication (Graham \& Faizo, 2017; Kardani et al., 2018; McBride, 2017). Early genes necessary for the initial genome amplification are expressed (E1, E2, E6 and E7). The URR, E1 and E2 are the primary genes required for successful replication. The E1 and E2 proteins are dependent on the host's replication machinery and DNA polymerase. In HR-HPVs, the E6 and E7 proteins have a pivotal role in driving cell proliferation in the basal layers and subverting pathways that signal cell growth arrest, thus supporting viral genome replication. In LR-HPVs, these proteins do not have a clear role in replication (Graham \& Faizo, 2017; Kardani et al., 2018; McBride, 2017). The viral genome doubles and segregates the viral DNA into daughter cells during basal cell division and replicates with the host DNA as episomes (S-phase). Low copy numbers (50-200 copies) of the viral episomes are maintained by maintenance replication to avoid host immune system detection (Graham \& Faizo, 2017; Kardani et al., 2018; McBride, 2017) (Figure 3).

In the third phase, the non-productive infectious state is maintained. Viral genome maintenance in proliferating basal cells follows a burst of viral DNA replication, also known as vegetative viral DNA amplification. The E4 and E5 proteins are expressed during this phase, although E4 is not essential for HPV11 replication (Graham \& Faizo, 2017; Kardani et al., 2018; McBride, 2017). Similarly, the late proteins (L1 and L2) necessary for viral assembly are expressed (Buck et al., 2013; Wang \& Roden, 2013). After differentiation of basal cells into keratinocytes, the cells exit the cell cycle (Graham \& Faizo, 2017; McBride, 2017). The purpose of the E6 and E7 proteins in HR- and LR-HPVs is to preserve a reservoir of infection and direct differentiating cells into the S-phase for viral amplification in the upper basal layers (Egawa \& Doorbar, 2017) (Figure 3).

In the last phase of viral replication, large numbers of progeny viral genomes are synthesised. Structural proteins expressed in the upper layers of epithelium and progeny virion particles are assembled and released upon cell death (Conway \& Meyers, 2009; Graham \& Faizo, 2017; Kardani et al., 2018;

McBride, 2017) (Figure 3). When HPV infects and replicates within the respiratory tract mucosa, papillomas may develop (Benedict \& Derkay, 2021; Hoesli et al., 2020; Seedat et al., 2013).


Figure 3: Human papillomavirus (HPV) replication in epithelial cells.
$H P V$ moves through the epidermis to promote infection. The proteins involved in the process include early proteins (E1, E2, E4, E5, E6 and E7) and late proteins (L1 and L2). Modified from Stanley, 2020.

### 1.6. Transmission of human papillomavirus

HPV is commonly transmitted during sexual contact with an infected partner, even when asymptomatic (Kardani et al., 2018). According to the Cancer Association of South Africa (CANSA), approximately $80 \%$ of people are infected with HPV at least once before the age of 50 years (https://cansa.org.za/). Globally, HPV is characterised as a prevailing sexually transmitted infection (STI). HPV can also be transmitted via auto-or hetero-inoculation, for example, kissing, non-penetrative sex, genital scratching, or finger-genital contact (Sabeena et al., 2017). In addition, HPV is also known to persist in a person's mouth, and HPV transmission between family members can thus be expected (Sabeena et al., 2017).

Although sexual transmission of HPV is predominant, HPV DNA has been identified in lesions in infants, young children, and women who are not yet sexually active, which suggests that the virus can be transmitted from an infected mother to child during or before childbirth (Kardani et al., 2018; Sabeena et al., 2017; Zahreddine et al., 2020; Zouridis et al., 2018). Infants born via caesarean section tend to have lower HPV transmission rates than those born naturally, although this delivery mode does not exclude the probability of infection (Tseng et al., 1998; Zouridis et al., 2018). Many well-conducted prospective studies have found that vertical transmission of HPV to newborn babies is relatively rare, with less than 5\% of children being infected with HPV from their mother (Smith et al., 2010; Watts et
al., 1998; Zouridis et al., 2018). However, results vary significantly, with other studies reporting higher (>20\%) rates of mother-to-infant transmission (Hahn et al., 2013; Rintala et al., 2005; Tseng et al., 1992, 1998). A systematic review of the literature published in 2018, including 421 HPV-positive mothers and their progenies, showed a $4.94 \%$ rate of vertical transmission of HPV and the relative risk of vertical transmission of HPV between women undergoing caesareans and vaginal deliveries to be 0.912 (Zouridis et al., 2018).

Few studies have reported on the differences between intrauterine transmission and transmission during labour and delivery. These studies produced varying results, with some studies unable to confirm intrauterine transmission and other studies reporting high rates of intrauterine HPV transmission. Nonetheless, an HPV positive mother continues to be a risk factor of infant infection. The prevalence of HPV infection in children infected in this manner is an essential area of research due to the impact on vaccination and management strategies.

### 1.7. Epidemiology of human papillomavirus type 11 in South Africa

A study conducted in the Western Cape between November 2012 to July 2013 on young women (age 16-24 years) determined that $71 \%$ of the population was infected with HPV, with half ( $50.1 \%$ ) of the HPV positive population infected with multiple HPV types. It was reported that between $1 \%$ and $3.1 \%$ of the HPV-positive population was infected with HPV11 and HPV6, respectively. These are the HPV types responsible for RRP development. This study also found that the younger women in the study population were more likely to test HPV positive than the older women ( $83 \%$ among ages $16-17$ years) (Giuliano et al., 2015).

A study that recruited sexually active, human immunodeficiency virus (HIV)-negative young women (age 16-22 years) from Cape Town and Soweto between November 2013 and December 2014 reported a lower HPV prevalence among the study population (66\%). In addition, this study revealed a lower percentage of infection with multiple HPV types (41.6\%). However, a higher prevalence of HPV11 and HPV6 infections ( $\sim 4 \%$ and $\sim 6 \%$, respectively) in the HPV-positive population was reported (Mbulawa et al., 2018). According to a previous surveillance study (2015-2016) for cervical HPV infections on unvaccinated women aged 18-20 years, it was determined that approximately two thirds ( $64.4 \%$ ) of the South African female population was infected with HPV, with a clear association between HPV and HIV. In addition, most women were infected with more than one HPV type with a median of three HPV types. Genotyping revealed that $3.4 \%$ of the population who tested HPV positive were infected with HPV11, and $4.8 \%$ were infected with HPV6 (Mbulawa et al., 2017).

A recent cross-sectional study showed that among women aged 38-55 years visiting a community health clinic in the OR Tambo district municipality of the Eastern Cape province of South Africa, HR-

HPV prevalence was $28.5 \%$. Of the 417 participants, $40.7 \%$ of women infected with an HR-HPV type were HIV-positive, and $14.4 \%$ of women reported having ulcers or warts in their lifetime. In addition, HIV-positive women with normal or abnormal cytology had a higher viral load than HIV-negative women (Taku et al., 2020). However, a limitation of most of these studies is that they targeted specific population groups or sampling areas and did not include children who predominantly suffer from RRP.

### 1.8. Human papillomavirus pathogenicity

HPVs are ubiquitous in the human population and often cause morbidity. LR-HPVs cause many seemingly benign lesions such as genital-, common-, and flat warts, as well as verrucas and other skin lesions. Generally, lesions caused by HPV are self-limiting and are frequently cleared by the cellmediated response (Sterling et al., 2001). However, HPV associated papillomas are often refractory to treatment and may persist. The development of cancer due to LR-HPV infection requires the host to be genetically susceptible or immunosuppressed to endure raised viral gene expression and recalcitrant lesions to progress to malignancy (Egawa et al., 2015; Georgescu et al., 2018; Pinidis et al., 2016; Reidy et al., 2004). Nevertheless, the association between LR-HPV and carcinoma remains ambiguous.

RRP, which is associated with HPV11, can persist for years regardless of surgical intervention, and can in some instances lead to malignant lesions in the lower respiratory tract and lungs (Gerein et al., 2005; Reidy et al., 2004). In these circumstances, the viral genome is integrated into the hosts' cell chromosome, signifying deregulated viral gene expression pursued by acquiring additional genetic and epigenetic modifications as seen in HR-HPV types (Huebbers et al., 2013; Reidy et al., 2004). In some RRP-associated cancers, rearrangement of the HPV11 genome has been observed, but viral gene expression patterns in this regard are still poorly understood. Compared to the prevalence of HPV6 and HPV11 infection in individuals, the development of RRP is rare, which may suggest that this disease should be considered as a multigene disease as HPV type and tissue-specific immune deficiency hinders the clearance and management of HPV6 and HPV11 infection. Therefore, in individuals who develop this disease, long term follow-up is essential (Bonagura et al., 2010; Seedat, 2020; Seedat et al., 2013).

HPV pathogenicity is determined by genotype, epithelial micro-environment, and the infection site and consequent pathogenesis is influenced mainly by viral protein function and regulation (Georgescu et al., 2018; Stanley, 2020). Knowledge of molecular pathogenesis at the gene regulation and protein function level is necessary to explain why LR-HPVs are less likely to cause progression to malignancy compared to HR-HPVs (Georgescu et al., 2018; Stanley, 2020). In tumour cells, the virus integrates into the host genome and interrupts the functions of the E 2 protein, which is responsible for E 6 and E 7 transcription, and subsequently results in the E6 and E7 oncoproteins being expressed at ele vated levels. As a result, infected cells are immortalised and unable to differentiate (Georgescu et al., 2018; Longworth \& Laimins, 2004; Stanley, 2020). In contrast to LR-HPV, HR-HPV E6 and E7 proteins
disturb epithelial differentiation and apoptosis by binding to host cell proteins with high affinity (Stanley, 2020). Epithelial differentiation and apoptosis disruption are stimulated through cellular proliferation, synthesis of viral DNA and interference in the cell cycle (Stanley, 2020).

In HR-HPVs, the E6 proteins form a complex with p53 and a ubiquitin ligase, E6-associated protein (E6AP), instigating cell cycle arrest and loss of apoptotic functions by blocking progression at the G1/Sphase checkpoint (Altamura et al., 2018; Pal \& Kundu, 2020). The E7 protein inactivates the pRB protein which disrupts its interaction with the E2 factor (E2F) at the G1-phase which subsequently results in disruption at the cell cycle control checkpoints (Liu et al., 2008; Pal \& Kundu, 2020). Additionally, the E7 protein interacts with the deacetylases, an enzyme, and the cyclin and cyclindependant kinase regulatory proteins to alter the cell cycle (Wang \& Roden, 2013). The interaction of E7 with enzymes triggers DNA synthesis and cell replication in generally inactive mature epithelial cells, subsequently stimulating differentiation-dependent viral DNA amplification, leading to pathological cell growth (Bedell et al., 1991; Pal \& Kundu, 2020). The virions then assemble and are released when the squamous cell layer is reached (Conway \& Meyers, 2009).

### 1.9. Host immune response to human papillomavirus

People with RRP often develop a robust serological response to the HPV vaccine; however, a serological response against natural infection is often delayed and produces low levels of antibodies. The delayed serological response is attributable to low levels of exposure to viral proteins due to immune evasion and the absence of viremia. It is postulated that only $50-70 \%$ of people infected with HPV develop specific antibodies (Buchinsky et al., 2020; Zahreddine et al., 2020).

Initially, patients diagnosed with RRP generate a measurable serum antibody response to HPV6 and HPV11 infection, indicating viral immune recognition. Therefore, immune dysfunction in patients diagnosed with RRP is indicative of HPV6 and HPV11 tolerance instead of a lack of viral recognition (Buchinsky et al., 2020; Ivancic et al., 2020). Furthermore, studies have demonstrated site-specific immune tolerance of HPV6 and HPV11 in the mucosa, suggesting that RRP is a multigene disease that polarises the immune system to tolerate local and chronic HPV6 and HPV11 infection (Buchinsky et al., 2020; Ivancic et al., 2020).

The adaptive immune system includes naïve T-lymphocytes that differentiate into cluster of differentiation (CD) 4+ and CD8+ cells (cell-mediated immunity) and B-lymphocytes, including plasma- and memory cells producing antibodies (humoral immunity). Phagocytes and antigen-specific cytotoxic T-lymphocytes, which form part of the cell-mediated immunity, are activated, and cytokines are released in response to the antigen. Due to cell-mediated immunity, the vast majority of papillomas regress within two years (Ivancic et al., 2020). B-lymphocytes, which display terminal differentiation
and plasma cells, named multiple myeloma oncogene-1 (MUM1+), produce antibodies that correlate with disease severity. HPV vaccination increases the level of antibodies and memory cells, thereby boosting immunity to HPV (Buchinsky et al., 2020; Ivancic et al., 2020).

Conversely, the innate immune system is a non-specific system that includes cells such as dendritic-, mast-, and natural killer cells, as well as macrophages and granulocytes. The innate immune system is activated by the presence of antigens, such as the L1 and L2 capsid proteins (Ivancic et al., 2020). The system further includes infiltration of leukocytes and production of nitric oxide, cytokines, and chemokines at the site of infection. Antibodies developed against HPV are associated with containment of the virus and papilloma regression (El Achkar et al., 2020; Ivancic et al., 2020).

Infection with HPV6 and HPV11 stimulates viral immune recognition and mounts a measurable serum antibody response (Ivancic et al., 2020; El Achkar et al., 2020). The CD3+ T-cells count represents the total number of T-lymphocytes. A higher number of CD3+ cells has been found in adults, who generally present with less severe RRP. A lower number of CD8+ cells was observed in juvenile patients, who generally present with more severe RRP. The immature juvenile immune system may affect the presentation of antigens and the secretion of pro-inflammatory cytokines, contributing to the severity of RRP. A higher number of CD8+ cells were detected in patients displaying low-grade dysplasia, whereas the CD4+ count remained unchanged in different degrees of dysplasia (El Achkar et al., 2020). Regulatory T-cells are responsible for immune self-tolerance by suppressing the activation of T-cells. A lower number of CD4+ regulatory T-cells, such as T-helper cells and natural killer cells, are observed in patients with frequent relapses (Ivancic et al., 2020; El Achkar et al., 2020).

A study describing the vertical transmission and clearance of immunoglobulin-G (IgG) antibodies against HPV6, HPV11, HPV16 and HPV18 in children showed that antibodies from newborns and mothers were moderately correlated, and that $80-100 \%$ of anti-HPV antibodies were cleared within the first two years of life in seropositive newborns, suggesting vertical transfer of the antibodies (Zahreddine et al., 2020). Thus, serological studies on HPV antibodies are essential to study natural immunity and to monitor the impact of HPV vaccination programs.

### 1.10. Recurrent respiratory papillomatosis

RRP is a debilitating disease characterised by the recurrent formation of benign papillomas, more commonly referred to as warts, in the mucosa of the respiratory tract (Benedict \& Derkay, 2021). The glottic and supraglottic regions of the larynx are most frequently affected. There are two forms of RRP, namely juvenile-onset RRP and adult-onset RRP (Novakovic et al., 2018; Seedat, 2020; Swain et al., 2020). HPV6 and HPV11 mainly cause RRP. However, HPV types 16, 18, 31, 33, 39, 44, 45, 55, and

70 have also been identified in respiratory papillomas (Hoesli et al., 2020; Peñaloza-Plascencia et al., 2000).

Papillomas in the airway cause dysfunction in the larynx and trachea and may lead to morbidity and possible mortality due to airway obstruction. Although the papillomas are histologically benign, the disease may be life-threatening to some individuals if treatment is not sought (Benedict \& Derkay, 2021; Seedat, 2020; Swain et al., 2020).

A study reviewing data from patients diagnosed with RRP between 2011 and 2015 in South Africa established that RRP in South Africa has a predominantly juvenile onset. This study reported the overall incidence of RRP to be 0.51 per 100000 population per year and the prevalence of RRP to be 1.39 per 100000 population. In children, the incidence of RRP was 1.34 per 100000 children per year and prevalence 3.88 per 100000 children. This study also found that RRP in children caused by HPV11 tended to be diagnosed at a younger age (median 3.2 years) than RRP caused by HPV6 (median 5.6 years) (Seedat \& Schall, 2018)

### 1.10.1. $\quad$ Risk factors associated with the development of recurrent respiratory papillomatosis

Numerous studies have compared the severity of RRP with the HPV genotype. It has been established that infection with HPV11 is possibly more severe, especially in the younger population, due to the rapid regrowth of the lesions (Intakorn \& Sonsuwan, 2014; Omland et al., 2014; Seedat, 2020). HPV infection disease is transmitted during childbirth or sexual contact with an infected partner (Kardani et al., 2018; Sabeena et al., 2017; Zouridis et al., 2018). Condylomas during pregnancy are regarded as a significant risk factor for acquiring juvenile-onset RRP by vertical HPV transmission (Rodier et al., 2013).

Risk factors for the development of RRP include any immunodeficiency or co-infection, particularly with herpes viruses and HIV. Prolonged exposure to the virus and a high viral load may also be risk factors for HPV11 infection and RRP development. Furthermore, the number of children born by a mother and the age at which birth is given may increase the likelihood of HPV11 infection and the development of associated diseases (Rodríguez-Álvarez et al., 2018; Taku et al., 2020). Another factor that may govern a patient's susceptibility to RRP includes genetic predisposition (Hahn et al., 2013; Rintala et al., 2005; Smith et al., 2010; Zouridis et al., 2018). However, more research on the vertical transmission of genes that increase the likelihood of HPV infection is needed. The tissue adjacent to the site of infection may also act as a latent virus reservoir. A trigger such as recuperation after surgical intervention to remove papillomas may prompt reactivation and replication of the HPV in the surrounding tissue (So et al., 2019).

Previous studies reported that RRP, although a benign neoplasm, might progress to carcinoma. Consequently, patients diagnosed with RRP showing evidence of dysplastic epithelium may be at possible increased risk of developing laryngeal cancer (Benedict \& Derkay, 2021; Cornall et al., 2013; Georgescu et al., 2018; Pinidis et al., 2016). However, a study on the association between LR-HPV types and the development of laryngeal neoplasia found no association between infection with HPV11 and laryngeal squamous intraepithelial neoplasia (SIN)-2+ or carcinoma. This study also found that coinfection with other HR-HPV types did not correlate with high-grade SIN and carcinoma. A significant preponderance of SIN2+ was instead identified in an HPV negative adult with adult-onset RRP (Omland et al., 2014).

### 1.10.2. Diagnosis of recurrent respiratory papillomatosis

The presenting symptoms of respiratory papillomas are progressive dysphonia, stridor, and respiratory distress. Laryngeal papillomas appear as off-white, exophytic, pedunculated, polypoidal masses. Single or multiple papillomas may be present (Ivancic et al., 2018; Seedat, 2020; Wilcox et al., 2014). Histopathological examination of a biopsy of the lesion gives a definitive diagnosis with RRP. Exophytic finger-like projections of keratinised squamous epithelium maintained by connective tissue and a vascular centre present in RRP. Vacuolated cells with noticeable cytoplasmic inclusions will be visible, indicating the presence of viruses (Ivancic et al., 2018; Seedat, 2020; Welschmeyer \& Berke, 2021; Wilcox et al., 2014).

### 1.10.3. Treatment of recurrent respiratory papillomatosis

Once a patient develops RRP, there is no definitive cure. Papillomas are removed using various techniques but tend to recur (Ivancic et al., 2018; Seedat, 2020; Swain et al., 2020). HPV DNA is present in uninvolved tissue and anatomical sites adjacent to the papillomas. Therefore, albeit with the removal of all papillomas, lesions tend to recur. The frequency of recurrence tends to decrease over time for most patients diagnosed with RRP, although this is not the case for all patients (Ivancic et al., 2018; Seedat, 2020; Seedat et al., 2013).

The mainstay of treatment for RRP is repeated microlaryngoscopic procedures aiming to clear the airway while preserving the mucosa and vocal folds (Ivancic et al., 2018; Seedat, 2020; Swain et al., 2020). However, in developing countries, cold steel instruments are often the only mean of removing papillomas (Seedat, 2020; Swain et al., 2020). Complications due to repeated surgical interventions may result in long-standing abnormal vocal quality. Patients with airway obstruction may also require a tracheostomy (Ivancic et al., 2018; Seedat, 2020; Swain et al., 2020).

A popular treatment option includes the microdebrider, which is quickly becoming the new gold standard for removing respiratory papillomas. The microdebrider causes minimal trauma to the
surrounding tissue and minimises the lower respiratory tract from being contaminated with papillomas and blood (Seedat, 2020; Swain et al., 2020). Controlled ablation is a relatively non-invasive low heat method used for the dissolution of soft tissue. This method is designed to cause minimal charring and burning of the tissue (Seedat, 2020; Swain et al., 2020).

Therapeutic vaccines that trigger the cell-mediated immune response are ideal for treating established HPV infections (Chabeda et al., 2018). Numerous approaches are being explored to develop prophylactic and therapeutic vaccines, including peptide-based vaccines, epitope-based vaccines, recombinant vaccines, bacteria-based vaccines, yeast-based vaccines, VLP-based vaccines, DNA vaccines, plant-based vaccines, dendritic cells (DC)-based vaccines, and protein-based subunit vaccines. Several candidates have progressed to clinical trials (Chabeda et al., 2018; Dadar et al., 2018). The E1, E2, E6 and E7 proteins are near-ideal targets for vaccines. The E6 and E7 oncoproteins are expressed constitutively and at high levels and have a low mutation rate to maintain malignancy. The E1 and E2 proteins are essential proteins expressed in all HPVs. These proteins are expressed at very high levels at the early stages of infection and are a valuable target for therapeutic vaccines. Currently, no therapeutic HPV vaccines have been approved for use in people infected with HPV, but promising vaccine candidates have been identified in numerous studies (Chabeda et al., 2018; Dadar et al., 2018).

Prophylactic vaccines have also been used as adjuvant therapy for RRP (Rosenberg et al., 2019). A systematic review and meta-analysis published in 2019 including patients previously diagnosed with RRP receiving the Gardasil® vaccine as a therapeutic option showed that the number of monthly surgical procedures for removal of laryngeal papillomas significantly reduced ( 0.35 to 0.06 per month) and that the intersurgical interval increased from seven months to over 34 months (Rosenberg et al., 2019).

Another adjuvant therapy for RRP treatment is intralesional cidofovir, an antiviral agent for treating cytomegalovirus retinitis in individuals diagnosed with acquired immunodeficiency syndrome (AIDS). The mechanisms of action of cidofovir involve decreasing the efficiency of DNA transcription of HPV. In RRP, cidofovir has been shown to increase relapse-free time and decrease the number of surgical procedures to remove papillomas (Fusconi et al., 2014; Gazia et al., 2020; Grasso et al., 2014; Graupp et al., 2013; Tjon Pian Gi et al., 2012; Welschmeyer \& Berke, 2021).

### 1.11. Prophylactic vaccines against human papillomavirus

There are currently three vaccines for protection against HPV: Cervarix ${ }^{\circledR}$, Gardasil ${ }^{\circledR}$, and Gardasil- $9^{\circledR}$. The vaccines are developed using VLPs and mimic a natural HPV infection. These VLPs are not infectious and do not contain any HPV DNA (Barra et al., 2019). The vaccines utilise the fact that HPV L1 proteins form VLPs, which are antigenically like native virions when expressed in a range of cell
types. The targeted HPV types are blocked from entering the host cell by eliciting neutralising antibodies that bind to the VLPs. However, these vaccines cannot eliminate established HPV infections since the target antigens are not expressed in infected cells (Benedict \& Derkay, 2021; Chabeda et al., 2018).

Cervarix is a bivalent vaccine that protects against HR-HPV types 16 and 18, responsible for causing high-grade cervical lesions. Gardasil is a quadrivalent vaccine that protects against two HR-HPV types, HPV16 and HPV18, and two LR-HPV types, HPV6 and HPV11, responsible for RRP development. Cervarix and Gardasil are aluminium-based first-generation vaccines proven to be highly immunogenic. Both vaccines exhibit cross-protection against HPV types 31, 33, 45, and 51 due to their phylogenetic relationship to HPV16 and HPV18. Gardasil-9 is a second-generation vaccine that protects against five HPV types in addition to the HPV types targeted by the quadrivalent Gardasil vaccine. These include HPV types 31, 33, 45, 52, and 58 (Barra et al., 2019; Chabeda et al., 2018; Dadar et al., 2018). The effectiveness of HPV vaccines in preventing infection by the targeted types has been documented in many countries (Brotherton, 2019).

Cervarix® has been incorporated into the national vaccination program by the South African National Department of Health in 2014 and is administered to girls aged nine to ten at public schools in two doses approximately six months apart. Approximately 354000 age-eligible girls in grade four have been vaccinated (Delany-Moretlwe et al., 2018). However, this vaccine does not confer immunity against other HPV types, such as HPV6 and HPV11. A study conducted by Mbulawa and colleagues in 2018 reported a high prevalence of the HPV types targeted by the Gardasil-9 vaccine ( $38.5 \%$ ) circulating amongst South-African women, encouraging the introduction of this vaccine into school health programs (Mbulawa et al., 2018). The introduction of this vaccine could be beneficial against the development of genital warts and RRP.

Nokavic and colleagues published a five-year report on surveys of 28 paediatric otolaryngologists on cases of RRP diagnosed post-HPV vaccination implementation in Australia. They demonstrated a decrease in the incidence of RRP from 0.16 to 0.02 per 100000 children and was the first study to report a decrease in the incidence of RRP following the implementation of an HPV vaccination program (Novakovic et al., 2018).

Future studies on the effectiveness of the 2014 HPV vaccination campaign are essential to inform vaccination campaigns, monitor the vaccination campaign's impact on circulating HPV types, and collect baseline data for future impact studies. Prevalence surveys reporting on the 2014 HPV vaccination campaign in South Africa are underway and expected to be published in 2024. Additionally, progress in immunology- and biotechnology-derived therapeutics, recombinant DNA technology and
molecular biology could multiply opportunities to fabricate vaccines that can effectively prevent and treat infection with HPV, which may in return result in a diminution of HPV cases globally.

### 1.12. Detection of human papillomavirus infections

Accurate identification and typing of HPVs relies on molecular biology techniques as the virus cannot be grown in conventional cell culture. At present, nucleic acid hybridisation -, signal amplification- and nucleic acid amplification assays are primarily used to identify and type HPV (Abreu et al., 2012).

### 1.12.1. Nucleic acid hybridisation assays

Nucleic acid hybridisation assays include Southern blotting, dot blot hybridisation and in situ hybridisation. However, nucleic acid hybridisation assays are not routinely used as they are timeconsuming and have low sensitivity (Abreu et al., 2012).

### 1.12.2. Signal amplification assays

Two signal amplification assays commonly utilised for diagnostic purposes are the Hybrid Capture ${ }^{\circledR} 2$ (hc2, Digene Corp., USA) and the Cervista® HPV HR assay (Abreu et al., 2012). The hc2 detects 13 HR-HPV types (HPV types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68). However, hc2 can cross-react with LR-HPV types, including HPV11 (Burd, 2016). The Cervista® HPV HR assay detects the same HR-HPV types as the hc2 plus HPV66 and has a lower cross-reactivity rate to other HPV types (Burd, 2016). Although signal amplification assays have increased sensitivity for genotyping and a low false-positive rate, they are not intended for single HPV genotyping (Abreu et al., 2012).

### 1.12.3. Nucleic acid amplification assays

There are numerous nucleic acid amplification-based methods to detect the presence of HPV in clinical samples reliably. Polymerase chain reaction (PCR) is a widely used, specific and sensitive way to detect the presence of HPV DNA and targets conserved regions such as the L1 capsid gene. In conventional PCR, consensus primers targeting the HPV L1 gene are often used for amplification. Primers, for example, MY09/MY11, GP5+/GP6+, PGMY and SPF10, are designed to detect various HPV types (Cornelissen et al., 1992; de Roda Husman et al., 1995). HPV genotypes can also be detected using primers targeting the E6 and E7 ORFs. The nucleotide sequences of the E6 and E7 ORFs contain many sequence variations (Bernard et al., 2006). Other ORFs such as the E5 ORF, in combination with the L1 ORF and URR, can also be used to distinguish between HPV types (Jelen et al., 2016). Following amplification, HPV can be genotyped using different techniques, for instance, genotype-specific primers, restriction-fragment length polymorphism (RFLP), sequencing, hybridisation, and linear probe assays (Abreu et al., 2012; Dixit et al., 2011). Real-time PCR is a specific and sensitive tool that has
been shown to reliably detect and genotype HPV. Real-time PCR targets a relatively small segment of the genome and can thus be used on archival clinical samples. This method delivers reproducible, rapid results and can be used on clinical samples (Abreu et al., 2012).

The Linear Array® HPV Genotyping Test (Roche Molecular Diagnostics, CA, USA) is an assay based on PCR amplification of an L1 gene segment and co-amplification of the $\beta$-globin gene, combined with reverse line blot hybridisation, which can distinguish between 15 HR-HPVs, three probable HR-HPVs, ten LR-HPVs (including HPV11) and nine other HPV genotypes with undetermined risk (Abreu et al., 2012). The Clinical arrays® HPV kit can be used to detect and genotype HPV and allows detection of 20 HR-HPV types and 15 LR types, including HPV11, and can also detect co-infections (Abreu et al., 2012). The Microplate colorimetric hybridisation assay (MCHA) is a PCR-based method that relies on amplifying a region within the L1 gene followed by colorimetric hybridisation to HPV type-specific probe micro-well plates to identify six different HR-HPV types (Abreu et al., 2012). The COBAS® 4800 HPV test combines automatic sample preparation and real-time PCR to detect 14 HR-HPV types (Burd, 2016). The CLART® Human papillomavirus 2 (Genomica, Madrid, Spain) is an amplificationbased methodology that targets a conserved region, L1, within the HPV genomic region and a region within the formylmethanofuran-tetrahydromethanopterin formyltransferase (FTR) gene. This test detects 35 HPV types using specific probes in a low-density microarray (Abreu et al., 2012).

### 1.12.4. Genotyping

Many HPV tests use probes that test for 13-14 of the most common HPV types. It has been proposed that testing for other HPV types may have a minimal effect on the knowledge of known, circulating HPV types. Still, there is an increasing need to distinguish between different HPV genotypes due to current advancements in the HPV field, evidence that certain HPV types correlate with high-grade carcinogenicity, the effect of multiple HPV infections on clinical presentation and the effect of vaccination campaigns on the current circulating HPV genotypes (Abreu et al., 2012; Brink et al., 2007; Godínez et al., 2014; Pinidis et al., 2016).

Traditionally, Sanger sequencing, also known as the chain termination method, was used for analysing DNA (Sanger et al., 1977). However, Sanger sequencing has limitations due to its restricted sensitivity, and it cannot perform a parallel investigation of multiple targets. Even though, Sanger sequencing remains the gold standard when analysing HPV DNA, developments in NGS technologies such as whole genome sequencing (WGS) and corresponding analytical tools have provided comprehensive novel genomic information on HPV and associated diseases on an unprecedented scale (Ambulos et al., 2016; Parker \& Chen, 2017; Tuna \& Amos, 2017). NGS enables the analysis of millions of DNA fragments simultaneously and has successfully been used to provide detailed information on the aetiology of HPV-driven diseases and cancers (Ambulos et al., 2016; Parker \& Chen, 2017; Tuna \&

Amos, 2017). Additionally, the advancement of NGS technology allows for the development of a highthroughput, affordable assay for HPV genotyping and has proven to be sensitive, specific, and capable of identifying multiple co-infected HPV genotypes from a single sample (Ambulos et al., 2016; Parker \& Chen, 2017; Tuna \& Amos, 2017).

HPV genome sequencing can be used to detect and genotype HPV on an assortment of clinical specimens such as blood, saliva, tissue, and formalin-fixed paraffin-embedded samples using the L1 ORF (Ambulos et al., 2016; Parker \& Chen, 2017). Additional sequencing is required to differentiate between the different lineages and sub-lineages (Burk et al., 2011; Maver et al., 2011). It has been determined that sequencing the E5 region and a small 208bp segment which includes the 3' end of E2 to the 5' end of NCR2, can be used to construct a phylogenetic tree representing the different lineages and sub-lineages of HPV11 (Jelen et al., 2016).

### 1.13. Phylogenetic analysis of human papillomavirus isolates

Phylogenetics is the study of phylogeny, which pertains to the evolutionary relationship among and within different organisms and species. Phylogenetics is linked to taxonomy, which is concerned with naming, describing, and classifying organisms (Balaban et al., 2019). Taxonomy relies on information gathered for phylogenetics during the classification of organisms. In phylogenetics, DNA sequencing methods such as Sanger sequencing or NGS investigate inheritable traits (Ambulos et al., 2016; Parker \& Chen, 2017; Sanger et al., 1977). This results in a schematic postulate of the relationship between organisms, termed a phylogenetic tree, that reflects the evolutionary history of said organism (Balaban et al., 2019; Hirose et al., 2018).

### 1.13.1. Sequence evolution

Before the division of any cell, its genome must be duplicated to be inherited by the offspring. Despite proof-reading mechanisms, errors still occur in the replication process. In time, genomic mutations such as point-mutations accumulate as traces of evolutionary variances. Point-mutations include nucleotide substitutions, which are the exchange of one nucleotide by another, the insertion of one or more bases, and the deletion of one or more bases. To distinguish between specific point-mutations, the sequence of a common ancestor is necessary to decide whether an insertion or deletion has occurred. Therefore, when investigating the evolutionary relationship of HPV11, a prototype sequence is used as a reference to identify nucleic heterogeneity (Balaban et al., 2019; Hirose et al., 2018).

Genomic mutations of DNA may also affect the amino acid sequence translated in protein-coding regions. Synonymous mutations occur if the nucleotide substitution does not alter the amino acid, and non-synonymous mutations occur when the nucleotide substitution changes the amino acid translated (Lebeuf-Taylor et al., 2019).

### 1.13.2. Multiple sequence alignment

Sequence alignment is an important research field in bioinformatics as it guides crucial tasks such as phylogenetic analysis. A multiple sequence alignment is often used to analyse homologous sequences. An alignment is a data matrix where homologous characters of the sequence are aligned in columns to identify discrepancies (Chowdhury et al., 2017).

### 1.13.3. Phylogenetic trees

Molecular Evolutionary Genetics Analysis (MEGA) is an example of software that provides a powerful and flexible interface for conducting phylogenetic analyses (Hall, 2013). Phylogenetic trees are a practical way to present the evolutionary relationship between organisms as the genetic sequence of contemporary sequences evolved from a common ancestor (Balaban et al., 2019). A phylogenetic tree consists of a root, nodes, tips and branches. The shape in which the branches connect the tips, nodes and root is termed the tree's topology. The tips of the phylogenetic tree represent the descendant or current taxa, and the nodes represent a common ancestor. The root represents the DNA sequence of all the species' sole common ancestor in the data set. Phylogenetic trees can be rooted or unrooted. A rooted tree denotes a common ancestor, whereas an unrooted tree has no origin and does not hypothesise the ancestral line (Balaban et al., 2019).

### 1.13.4. Types of phylogenetic methods

Methods of tree reconstruction based on DNA sequences include maximum parsimony methods, evolutionary distances between sequences and approaches applying the maximum likelihood principle (Rusinko \& McPartlon, 2017; Serdoz et al., 2017).

Neighbor-Joining is a distance-based agglomerative clustering technique that takes a distance matrix detailing the distance between each pair of taxa as an input. Distance-based methods alter the sequence data into a pairwise similarity matrix to use during tree interpretation. The algorithm starts with a completely unresolved tree and iterates until the tree is resolved and branch lengths are established (Rusinko \& McPartlon, 2017). The Neighbor-Joining method is suited for complex datasets comprising lineages with variable rates of evolution and permits correction for numerous substitutions. Additionally, the Neighbor-Joining method is suited for large input data sets and generates results in a small amount of time. However, due to the Neighbor-Joining tree being distance matrix based, the output information is often limited, and only one out of several possible trees are depicted depending on the model of evolution used (Kuhner \& Felsenstein, 1994; Saitou \& Nei, 1987).

Parsimony reconstruction and maximum likelihood methods are both character-based methods that use all known evolutionary information, such as nucleotide substitutions, to identify the most likely
ancestral relationship (Serdoz et al., 2017). Character-based methods aim to create an algorithm for scoring the probability that a given tree would produce the observed sequences at its leaves. The algorithm then filters through the possible trees to identify a tree with the maximum probability of producing the correct results (Serdoz et al., 2017).

The most popular character-based method when constructing a phylogenetic tree for HPV11 is the maximum likelihood method (Burk et al., 2011; Jelen et al., 2016). This algorithm provides probabilities that a particular evolutionary model has generated the observed data. This algorithm uses each position in a sequence alignment and considers all possible trees. The higher the probability of the sequence given, the more the tree is favoured until the maximum likelihood is determined. Since an evolutionary model can be chosen, this tree is suitable for a large variety of groups (Miguel Rocha \& Ferreira, 2018; Serdoz et al., 2017). The supposed advantages of maximum likelihood methods include a lower variance rate compared to other methods, and it is rarely affected by errors during sampling. The method tends to be robust to several violations of evolutionary model assumptions, and even with short sequences, it outperforms alternate methods. However, the maximum likelihood method is computationally intensive, and the result is dependent on the model of evolution used (Miguel Rocha \& Ferreira, 2018; Serdoz et al., 2017).

### 1.14. Problem statement

During a previous study conducted at the Division of Virology, UFS, a novel HPV11 lineage B isolate was identified, warranting investigation of further isolates for clarification of lineages currently in circulation (Makatsa, 2012). It is essential to characterise the current strains circulating amongst the community, compare the various lineages and sub-lineages with disease severity in future studies, and monitor the impact of the vaccination campaign on the circulating HPV types and guide vaccine development. Baseline data on circulating HPV genotypes may effectively monitor the impact of the vaccination program on the community. Also, determining the genotype of HPV responsible for RRP in patients may have prognostic implications. Hence, this study investigated HPV11 lineages circulating in patients with RRP and identify novel lineages.

### 1.15. Aims and objectives

The aim of this study is to genetically characterise HPV11 isolates from patients diagnosed with RRP treated at Universitas Academic Hospital, Free State, South Africa.

To achieve these aims the following objectives were identified:

- To determine the nucleotide sequence of the L1, URR, E5a/b and a segment of E2 genomic regions of HPV11 isolates from patients with RRP in the Free State, South Africa.
- To conduct phylogenetic analyses using sequence data from HPV11 isolates to determine the genetic relationship and to determine the presence of HPV11 variants and lineages.
- To generate and analyse the whole genome sequences of novel HPV11 variants.


# CHAPTER 2- Phylogenetic analysis of human papillomavirus type 11 isolates 

### 2.1. Introduction

HPV11 is a causative agent of RRP, a benign laryngeal neoplasm presenting mainly in children (Benedict \& Derkay, 2021). To date, no published data on the HPV11 intratypic variants circulating in the Free State region is available. Therefore, phylogenetic analysis may aid in identifying circulating variants in patients with RRP.

It is commonly accepted that natural selection influenced the evolution of organisms and their contemporary genes. An increasing number of DNA sequences are being analysed today, and the study of the ancestry of organisms and their phylogenetic trees is in high demand. A better understanding of phylogenetic trees and the evolutionary processes that they model is necessary to gain insight into how organisms have mutated.

Phylogenetic analysis investigates evolutionary development in different species by analysing genome information such as DNA, deepening our knowledge of how different species' genomes develop (Charleston, 2013; Felsenstein, 1981; Hall, 2013). To date, over 220 HPV types have been identified and classified into different risk groups through phylogenetic analysis, according to The International HPV Reference Centre (Bernard et al., 2010; de Villiers et al., 2004) (https://www.hpvcenter.se/human_reference_clones). Although Sanger sequencing remains the gold standard for sequencing, especially when sequencing single genes, recent advances in NGS technology has led to NGS being applied as the preferential technology for several kinds of molecular studies and typing (Ambulos et al., 2016; Kocjan et al., 2015).

Analysis of sequence heterogeneity across the complete genome of HPV is frequently used to identify HPV variants. The ORF nucleotide sequence coding for the L1 capsid protein is sufficient for classifying HPV types and is used to represent the whole genome variation due to its high level of conservation (Burk et al., 2011, 2013). However, there is a loss of clear resolution when the L1 sequence is used to identify intratypic HPV variants. To discriminate between different HPV11 lineages, the URR and a coding region such as the L1 ORF are necessary, but for clustering of sub-lineages, additional sequencing is required (Burk et al., 2011, 2013). Sequence data can be aligned to a prototype genome to identify nucleotide heterogeneity, and variant lineages and sub-lineages of HPV11 can be identified by the percentage of nucleotide differences in the complete genome. A nucleic acid change of $0.5 \%$ $1.0 \%$ across the whole genome defines a variant sub-lineage, $1 \%-10 \%$ defines a variant lineage, and $>10 \%$ defines a novel HPV type (Burk et al., 2011). In 2011, the circulating HPV11 variants were
classified as A1 and A2, and in 2016, whole-genome analysis of 78 HPV11 complete genomes revealed two additional sub-lineages (A3 and A4) and a novel lineage B (Burk et al., 2011; Jelen et al., 2016). Nucleotide variances in genomic DNA can lead to changes in translated gene products. Therefore, the amino acid sequence can also help study the evolutionary relationship between organisms (Charleston, 2013; Felsenstein, 1981; Hall, 2013).

Phylogenetic trees are constructed by use of sequence data and are helpful to depict the evolutionary relationships between organisms. Phylogenetic trees that reflect the evolutionary history of an organism is preferred (Charleston, 2013; Felsenstein, 1981; Hall, 2013). Finding an accurate phylogenetic tree for a set of related species with only DNA is complex. Statistical inference using statistical models and Markov models has helped scientists estimate how comparable a phylogenetic tree is to the actual phylogenetic tree for a specified set of DNA sequences (Charleston, 2013; Felsenstein, 1981; Hall, 2013).

Neighbor-Joining is a distance-based agglomerative clustering technique that takes a distance matrix detailing the distance between each pair of taxa as an input (Rusinko \& McPartlon, 2017). The Neighbor-Joining method is suited for large, complex datasets comprising lineages with variable rates of evolution and permits correction for numerous substitutions and generates results in a small amount of time (Kuhner \& Felsenstein, 1994; Saitou \& Nei, 1987).

However, maximum likelihood trees are preferred when constructing a tree for HPV11 isolates and use an explicit evolutionary model. This method provides probabilities that a particular evolutionary model will generate the observed data until the tree with the maximum likelihood is determined (Charleston, 2013; Felsenstein, 1981; Hall, 2013). The method has been proven to outperform alternate methods (Charleston, 2013; Felsenstein, 1981; Hall, 2013).

The current study explored phylogenetic relatedness and nucleotide variability among HPV11 isolates from patients diagnosed with RRP. The L1, URR, E5a/b and a segment of the E2 ORF were targeted for sequencing. The concatenated L1-URR is necessary to discriminate between HPV11 lineages (Burk et al., 2011), and according to a study published in 2016, the concatenated L1-URR-E5a/b can discriminate between different lineages and sub-lineages (Jelen et al., 2016). The 208bp region in the E2 ORF has also been suggested to represent whole-genome variation and could therefore be used to construct phylogenetic trees representative of the different lineages and sub-lineages (Jelen et al., 2016).

Characterising the current circulating strains will aid in future disease severity studies, guide vaccine development and have prognostic implications. Hence, this study investigated HPV11 lineages and sublineages circulating in patients with RRP.

### 2.2. Materials and methods

### 2.2.1 Study samples

During a previous study conducted at the Division of Virology at the University of the Free State, Bloemfontein, 94 laryngeal papilloma biopsies were collected from 2008 to 2018 and stored at $-80^{\circ} \mathrm{C}$ (ETOVS 194/2007 and ECUFS 6/2011) (Appendix 1). Informed consent for collection and storage of biopsies was obtained from each patient by Professor Seedat from the Department of Otorhinolaryngology, Faculty of the Health Sciences, University of the Free State.

All samples were assigned a VBD number followed by the year of collection. DNA was extracted from each biopsy by use of the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions and stored at $-20^{\circ} \mathrm{C}$. Samples were screened for HPV DNA by PCR, targeting a region of the L1 ORF using the GoTaq ® Flexi DNA Polymerase mediated PCR amplification kit (Promega, Madison, USA) according to manufacturer's instructions. Following PCR samples were cleaned with the Promega Wizard ${ }^{\circledR}$ SV Gel PCR Clean-up System kit according to manufacturer's instructions (Promega, Madison, USA). The positive reactors were genotyped by nucleotide sequence determination of the amplicon with the BigDye ® Terminator v 3.1 Cycle Sequencing kit (Thermo Fisher Scientific, Illinois, USA) and aligned with sequence data retrieved from GenBank using Basic Local Alignment Search Tool (BLAST) analysis. A novel HPV11 lineage B isolate identified in a previous study was also included in the current study (Makatsa, 2012).

### 2.2.2 Study setting and population

All tests were performed at the Division of Virology, Faculty of Health Sciences at the University of the Free State (UFS). The study population included patients diagnosed with RRP at the Universitas Academic Hospital, Bloemfontein, a referral hospital including patients mainly from the Free State province, but also from surrounding regions. Eighty-one isolates from 60 patients were available for this study. Additional sequence data for 28 geographically distinct representatives of each HPV11 lineage, including lineages A and sub-lineages A1 to A4, and lineage B, were retrieved from GenBank (Appendix 2). The current study was approved by the Health Sciences Research Ethics Committee of the University of the Free State (UFS-HSD2019/1109/2708) (Appendix 3).

### 2.2.3 Amplification of five human papillomavirus genomic regions

PCR amplification of the complete L1 ORF, URR, E5a ORF and E5b ORF was performed with primers as described by Maver and colleagues (2011) summarised in Table 1. Primers flanking a segment of the E2 were designed according to whole genome HPV11 sequences retrieved from GenBank (Appendix 4). Sequence data for the E2 ORF from 28 HPV11 isolates from different geographic distribution were
retrieved from GenBank and aligned using ClustalX v2.1 (Larkin et al., 2007). Primers were designed to amplify the $3^{\prime}$ of E2 to the $5^{\prime}$ of the NCR2 which is 208bp.

The primers' specificity was determined with BLAST analysis and are summarised in Appendix 5. The GC content and melting temperature (Tm) of the forward primer (HPV11-E2seg-F) and reverse primer (HPV11-E2seg-R) were determined with Thermo Fisher Scientific Tm calculator (Allawi \& SantaLucia, 1997) and are summarised in Table 1. PCR amplification of the partial E2 ORF was performed with the designed primer pair which targets a region of approximately 331bp, flanking the target partial E2 ORF. Amplification of all samples was performed with the Applied Biosystems ${ }^{\mathrm{TM}}$ Proflex ${ }^{\text {TM }}$ PCR system (Thermo Fisher Scientific, Illinois, USA). PCR amplification was performed using the Go Taq®G2 Hot Start Polymerase (Promega, Madison, USA) according to the manufacturer's instructions.

Briefly, all reaction mixtures were prepared in $200 \mu \mathrm{l}$ PCR reaction tubes. Each reaction comprised $4 \mu \mathrm{l}$ template HPV11 DNA, $10 \mu \mathrm{l}$ 5GGreen Go Taq ${ }^{\circledR}$ flexi buffer, $2 \mathrm{mM} \mathrm{MgCl}{ }_{2}$ solution, 0.2 mM PCR nucleotide mix, $0.4 \mu \mathrm{M}$ forward primer, $0.4 \mu \mathrm{M}$ reverse primer, 1.25 U Go Taq ${ }^{\circledR} \mathrm{G} 2$ Hot Start Polymerase and nuclease free water up to $50 \mu$. A no template control was included in each PCR amplification run by replacing the template with nuclease free water during preparation of the reaction. PCR components are included in Table 2.

PCR amplification conditions were as follows: initial denaturation at $95^{\circ} \mathrm{C}$ for two minutes for one cycle, followed by 30 cycles of denaturation at $95^{\circ} \mathrm{C}$ for 30 seconds, annealing at $47^{\circ} \mathrm{C}$ for 30 seconds and elongation at $72^{\circ} \mathrm{C}$ for two minutes. Lastly, final elongation at $72^{\circ} \mathrm{C}$ for five minutes before cooling to $4^{\circ} \mathrm{C}$ indefinitely.

Table 1: List of primers used for sequencing the L1, E5a/b, upper regulatory region (URR) and a segment of the E2 of human papillomavirus type 11 (HPV11). Primers HPV11-L1F, HPV11-L1R, HPV11-LCR-F, HPV11-LCR-R, HPV11-E5F and HPV11-E5R are modified from Maver et al., 2011. Primers HPV11-E2seg-F and HPV11-E2seg-R were designed according to whole genome HPV11 sequences retrieved from GenBank.

| Region | Primer name | Primer sequence $\left(5^{\prime} \rightarrow 3^{\prime}\right)$ | $\begin{gathered} \mathrm{GC} \\ \text { content }{ }^{\mathrm{e}} \end{gathered}$ | Tm ${ }^{\text {e }}$ | Primer binding $\text { site }{ }^{d}$ | Gene size ${ }^{\text {f }}$ | Amplicon size ${ }^{\text {f }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| L1 | HPV11-L1F ${ }^{\text {a }}$ | TGTTTTTATTACAG <br> GTTCTGACTTC | 32.00\% | $56.0{ }^{\circ} \mathrm{C}$ | nt 5682-5706 | 1,506bp | 1,770bp |
|  | HPV11-L1R ${ }^{\text {a }}$ | AAAACATACATAC ACATTCCACAAA | 28.00\% | $55.4{ }^{\circ} \mathrm{C}$ | nt 7451-7427 |  |  |
| URR | HPV11-LCR-F ${ }^{\text {a }}$ | CTACAGCCCCCAA <br> ACGAA | 55.56\% | $57.5{ }^{\circ} \mathrm{C}$ | nt 7236-7253 | 756bp | 823bp |
|  | HPV11-LCR-R ${ }^{\text {a }}$ | CGTGGAGGCATCTT TACTTTC | 47.62\% | $58.5{ }^{\circ} \mathrm{C}$ | nt 125-105 |  |  |
| E5a/b | HPV11-E5F ${ }^{\text {a }}$ | AATACCACCCACC ATTAGGC | $50.00 \%$ | $56.4{ }^{\circ} \mathrm{C}$ | nt 3772-3791 | $500 \mathrm{bp}{ }^{\text {g }}$ | 684bp |
|  | HPV11-E5R ${ }^{\text {b }}$ | $\begin{aligned} & \text { GGCTGACGCACGTT } \\ & \text { TAC } \end{aligned}$ | 58.82 \% | $56.2{ }^{\circ} \mathrm{C}$ | nt 4455-4439 |  |  |
| E2 segment ${ }^{\text {c }}$ | HPV11-E2seg-F ${ }^{\text {c }}$ | ACAGTGCAGCTAC GCCTATA | 50,00\% | $56.0{ }^{\circ} \mathrm{C}$ | nt 3561-3580 | 208bp | 331bp |
|  | HPV11-E2seg-R ${ }^{\text {c }}$ | TTGTACAGGCACTA CCTCCATAC | 47.83\% | $60.1{ }^{\circ} \mathrm{C}$ | nt 3891-3869 |  |  |
| ${ }^{\text {a }}$ Primers reported by Maver et al. , 2011. |  |  |  |  |  |  |  |
| ${ }^{\text {c }}$ Primers designed according to geographically distinct whole genome human papillomavirus type 11 sequences retrieved from GenBank (GenBank Acc. No M14119.1; JN644141.1; JQ773408.1; MN788368.1; LN833187.1). |  |  |  |  |  |  |  |
| ${ }^{\text {d }}$ Positions of the nucleotides determined with respect to the prototype human papillomavirus type 11 genome (GenBank acc. no. M14119.1). |  |  |  |  |  |  |  |
| ${ }^{\mathrm{e}}$ Melting temperature (Tm) and GC content determined by Thermo Fisher Scientific Tm calculator (Allawi \& SantaLucia, 1997). |  |  |  |  |  |  |  |
| ${ }^{\mathrm{f}}$ Size determined with respect to the prototype human papillomavirus type 11 genome (GenBank acc. no. M14119.1). |  |  |  |  |  |  |  |

### 2.2.4 Agarose gel electrophoresis

The size of the PCR amplicons generated during PCR amplifications were determined by separating $5 \mu 1$ aliquots of PCR products with gel electrophoresis. A $1 \%$ agarose gel was used to determine the amplicon size and integrity of the amplicon generated by the PCR. A table summarising the preparation of agarose gels is included in Appendix 7. For all amplicons, a $1 \%$ gel was prepared with 1 g Seakem®

LE agarose powder and 100 ml of 1 x Tris-Acetic acid- (ethylenediaminetetraacetic acid)- EDTA (TAE) buffer at pH 8.0 and the gel was electrophoresed at 90 volts $(\mathrm{V})$ for 45 minutes. The preparation of TAE is included in Appendix 6. The agarose gel was stained with a GelRed nucleic acid gel stain (Thermo Fisher Scientific, Illinois, USA) (Appendix 8) and visualised with the BioRad Molecular Imager Gel Doc ${ }^{\text {TM }} \mathrm{XR}+$ with Image Lab ${ }^{\text {TM }}$ Software to determine the fragment sizes according to the known DNA size marker (BioRad, California, USA). The O'GeneRuler ${ }^{\text {TM }}$ 100bp DNA ladder SM1173 and SM0333 (Fermentas, Illinois, USA) containing DNA fragments from 100 to 10000 bp was used to estimate the size of the amplicons. Expected amplicon sizes varied, with the primers targeting the L1 ORF producing the largest amplicon (1770bp), followed by the primers targeting the URR (823bp), the E5a/b ORF (685bp) and the E2 segment (331bp).

Table 2: Polymerase chain reaction (PCR) components for amplifying human papillomavirus type 11 (HPV11) L1 gene, E5a/b gene, and upper regulatory region (URR).

| PCR components | Volume | Final concentration |
| :--- | :---: | :--- |
| 5 XGreen Go Taq® flexi buffer | $10 \mu \mathrm{l}$ | 1 x |
| MgCl2 solution, 25 mM | $4 \mu \mathrm{l}$ | 2 mM |
| PCR nucleotide mix, 10 mM each | $1 \mu \mathrm{l}$ | 0.2 mM |
| Forward primer: HPV11-L1F $(20 \mathrm{pmol} / \mu \mathrm{l})$ | $1 \mu \mathrm{l}$ | $0.4 \mu \mathrm{M}$ |
| Reverse primer: HPV11-L1R $(20 \mathrm{pmol} / \mu \mathrm{l})$ | $1 \mu \mathrm{l}$ | $0.4 \mu \mathrm{M}$ |
| GoTaq ®G2 Hot Start Polymerase $(5 \mathrm{U} / \mu \mathrm{l})$ | $0.25 \mu \mathrm{l}$ | 1.25 U |
| Template DNA | $4 \mu \mathrm{l}$ | $\mathrm{N} / \mathrm{A}$ |
| Nuclease free water | $28.75 \mu \mathrm{l}$ | $\mathrm{N} / \mathrm{A}$ |
| Total | $50 \mu \mathrm{l}$ | $\mathrm{N} / \mathrm{A}$ |

### 2.2.5 Purification of PCR product

Following electrophoresis, DNA amplicons were purified using the GeneJET PCR Purification Kit (Thermo Scientific, Illinois, USA) according to the manufacturer's instructions. Briefly, equal amounts of binding buffer and PCR mixture were mixed thoroughly and transferred to the purification column before centrifuging for 60 seconds on the Eins-Sci E-C15-12.2 High Speed Micro Centrifuge (Eins Sci, South Africa) at 25200 relative centrifugal force (rcf) and discarding the flow-through. A $700 \mu 1$ aliquot of wash buffer was added to the purification column and centrifuged twice for 60 seconds (2520xrcf) and discarding the flow-through twice. The column was transferred to a $1,5 \mathrm{ml}$ microcentrifuge tube and
between $30 \mu \mathrm{l}$ to $50 \mu \mathrm{l}$ aliquots of elution buffer was added to the column. The tube was centrifuged at $2520 x r c f$ for 60 seconds and the column was discarded. The eluted DNA was stored at $-20^{\circ} \mathrm{C}$ until use.

### 2.2.6 Determination of DNA concentration

DNA concentration was determined with the NANODROP 2000 spectrophotometer (Thermo Fisher Scientific, Illinois, USA).

### 2.2.7 Sequencing

The nucleotide sequences of each amplicon were determined with Sanger sequencing to identify the nucleotide composition in order to identify the variant lineages and sub-lineages. Sequencing reactions were prepared with the BigDye Terminator v3.1 Sequencing Reaction kit according to the manufacturer's instructions (Thermo Fisher Scientific, Illinois, USA). Briefly, the purified HPV isolates were diluted with nuclease-free water to $20 \mathrm{ng} / \mu$ l. Sequencing primers were diluted with nuclease-free water to a concentration of $3.2 \mathrm{pmol} / \mu 1$.

For the L1 ORF, four primers were used for sequencing as described by Maver and colleagues in 2011 (Table 3). For the URR, primers used for the initial PCR as described in Table 1, with one additional primer (HPV11-LCR-FF, 5'- TTCGGTTGCCCTTACATACA -3', nt7598-7617) were used as described by Maver and colleagues (2011). For E5a/b ORF and the E2 segment, primers used for the initial PCR as described in Table 1 were used. Each amplicon was sequenced using bidirectional sequencing.

Table 3: List of primers used for sequencing the L1 gene of human papillomavirus type 11 (HPV11) from Maver et al., 2011.

| Region | Primer name | Primer sequence $\left(5^{\prime} \rightarrow 3^{\prime}\right)$ | GC content ${ }^{\text {c }}$ | Tm ${ }^{\text {c }}$ | Primer binding site ${ }^{\text {b }}$ | Gene size ${ }^{\text {d }}$ | Amplicon size ${ }^{\text {d }}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| L1 | HPV-11-L1U1 ${ }^{\text {a }}$ | CGTAAACGTATTCC CTTATTTTTT | 29.17\% | $54.24^{\circ} \mathrm{C}$ | nt 5743-5766 | 1,506bp | 871bp |
|  | HPV-11-L1U1R ${ }^{\text {a }}$ | TGATCTGTTATTAC CCCCTTTTAC | 37.50\% | $56.16^{\circ} \mathrm{C}$ | nt 6613-6590 |  |  |
|  | HPV-11-L1U2 ${ }^{\text {a }}$ | TTTTATTTGCGAAA GGAACA | 30.00\% | $51.86{ }^{\circ} \mathrm{C}$ | nt 6503-6522 |  | 809bp |
|  | HPV-11-L1U2R ${ }^{\text {a }}$ | ACATATAAATAAC ACAACACACTGAC | 30.77\% | $56.30^{\circ} \mathrm{C}$ | nt 7311-7286 |  |  |
| ${ }^{a}$ Primers re <br> ${ }^{\mathrm{b}}$ Positions <br> ${ }^{c}$ Melting te <br> ${ }^{\mathrm{d}}$ Size detern | ted by Maver et al., <br> he nucleotides detern erature (Tm) and GC <br> ed with respect to the | d with respect to the prototyp ent determined by Thermo totype human papillomaviru | pe human papill <br> sher Scientific T <br> type 11 genom | rus type 1 <br> ulator (A <br> Bank acc | me (GenBank acc. no. M SantaLucia, 1997). 14119.1). | 119.1). |  |

A reaction mixture for every primer containing 2 ul 5 x sequencing buffer, 1 ul BigDye v3.1 ready reaction, 1 ul primer $(3,2 \mathrm{pmol} / \mu \mathrm{l})$ was prepared. A total volume of $2 \mu$ l diluted HPV DNA and $8 \mu \mathrm{l}$ of each reaction mixture was added to the well of the 96 -well plate. The sequencing components for nucleotide determination of HPV11 isolates are included in Table 4. The plate was sealed with thermoadhesive film and briefly centrifuged. The plate was placed in the Applied Biosystems ${ }^{\text {TM }}$ Proflex ${ }^{\text {TM }}$ PCR system (Thermo Fisher Scientific, Illinois, USA) for DNA amplification. PCR amplification conditions were as followed: initial denaturation at $96^{\circ} \mathrm{C}$ for one minute and one cycle, followed by 25 cycles of denaturation at $96^{\circ} \mathrm{C}$ for 10 seconds, annealing at $50^{\circ} \mathrm{C}$ for five seconds and elongation at $60^{\circ} \mathrm{C}$ for four minutes before cooling to $4^{\circ} \mathrm{C}$ until further use.

Table 4: Sequencing reaction components for nucleotide determination of human papillomavirus type 11 (HPV11) isolates.

| Components | Volume |
| :--- | :---: |
| $5 \times$ Sequencing Buffer | $2 \mu \mathrm{l}$ |
| Primer $(3,2 \mathrm{pmol} / \mu \mathrm{l})$ | $1 \mu \mathrm{l}$ |
| BigDye v 3.1 | $1 \mu \mathrm{l}$ |
| Template DNA | $2 \mu \mathrm{l}$ |
| Nuclease free water | $4 \mu \mathrm{l}$ |
| Total | $10 \mu \mathrm{l}$ |

Following DNA amplification, ethanol/EDTA precipitation was performed. Nuclease free water and 125 mM EDTA was combined in a $2: 1$ ratio, and a $15 \mu \mathrm{l}$ aliquot of the diluted EDTA solution was added to each well. Thereafter, a $60 \mu \mathrm{l}$ aliquot of cold $96-100 \%$ ethanol stored at $-20^{\circ} \mathrm{C}$ was added to each well. A silicone cover was placed on the 96 -well plate before briefly vortexing the plate. The plate was placed in the $-20^{\circ} \mathrm{C}$ freezer for five minutes. Afterwards, the plate was centrifuged at 2720 xrcf for 80 minutes at $4^{\circ} \mathrm{C}$. After centrifugation, the supernatant was aspirated entirely without disturbing the pellet. A volume of $200 \mu \mathrm{l}$ cold $70 \%$ ethanol stored at $-20^{\circ} \mathrm{C}$ was added to each well. The plate was sealed and centrifuged at 2720 xrcf for 40 minutes at $4^{\circ} \mathrm{C}$. The liquid was aspirated without disturbing the pellet. The plate was then placed in the Applied biosystems ${ }^{\text {TM }}$ Proflex ${ }^{\text {TM }}$ PCR system (Thermo Fisher Scientific, Illinois, USA) at $94^{\circ} \mathrm{C}$ for $25-40$ seconds to evaporate ethanol. Lastly, a $10 \mu 1$ aliquot $\mathrm{Hi}-\mathrm{Di}$ was added to each well, and the samples were denatured at $95^{\circ} \mathrm{C}$ for five minutes in the Applied biosystems ${ }^{\text {TM }}$ Proflex ${ }^{\text {TM }}$ PCR system (Thermo Fisher Scientific, Illinois, USA).

The Applied Biosystems ${ }^{\text {TM }} 3500 \mathrm{Xl}$ Genetic Analyzer (Thermo Fisher Scientific, Illinois, USA) was used to sequence all samples. Briefly, a primer, enzyme and four fluorescently labelled dideoxynucleotides triphosphates (ddNTPs) are added to the reaction. The ddNTPs emit light at different wavelengths when excited by a laser. This emission can be captured by a camera and
converted to a chromatogram. As the fluorescently labelled extension products pass the laser, each nucleotide is "called". Raw sequence data were analysed as described below.

### 2.2.8 Data analysis

## Analysis of nucleotide variation

Sequence data for five regions, namely the L1- (1506bp), URR- (758bp), E5a- (276bp), E5b region (225bp) and a section from the 3 ' of E2 to the 5' of the NCR2 (208bp) was determined. The E5a/b region contains one overlapping nucleotide (E5a - nt3871-4146; E5b - nt4146-4370) and therefore consist of 500bp. Raw sequence data was manually verified with Unipro UGENE (Okonechnikov et al., 2012). Briefly, the ends of the reads were trimmed based on the quality value. Each region was sequenced bi-directionally, and the chromatograms of the filtered trimmed reads were aligned and inspected to correct for possible nucleotide-calling faults due to ambiguous peaks.

Genomic variants and genomic positions in HPV11 L1-, E5a, E5b- and a segment of the E2 ORF and the URR were identified by comparing the sequence data with the prototype HPV11 genome (GenBank accession number M14119.1). For the URR, a corrected sequence with a $2 b p$ insertion at genomic position 7717-7718 was used when determining genomic variants and genomic positions (Maver et al., 2011). The nucleotide substitutions, nucleotide insertions, and nucleotide deletions were identified for the isolates sequenced in this study with respect to prototype strain (GenBank accession number M14119.1) by use of MEGA X software (Kumar et al., 2018).

To estimate the percentage of variation within coding regions, the number of nucleotide variations compared to the A1 prototype sequence M14119.1 were converted into a percentage using the following equation:

$$
\frac{\text { Number of nucleotide variances within a spesific coding region compared to M14119.1 }}{\text { Number of nucleotides within spesific coding region of M14119.1 }} \times 100
$$

## Degree of divergence

The degree of divergence was calculated using pairwise analysis of percentage divergence of nucleotides with MEGA X software using concatenated E5a/b-L1-URR sequences of sequences obtained in the current study and sequences retrieved from GenBank. Codon positions included were $1^{\text {st }}, 2^{\text {nd }}, 3^{\text {rd }}$, and non-coding. All positions with less than $95 \%$ site coverage were removed using the pairwise deletion option.

## Analysis of amino acid variation

The amino acid data order was determined for the L1- (501aa), E5a- (91aa), and E5b region (74aa) with use of a codon chart included in Appendix 9. Amino acid variants in HPV11 L1-, E5a, E5b ORFs were identified by comparing the amino acid data with the prototype HPV11 (GenBank accession number M14119.1).

## Phylogenetic analysis

All sequences were aligned with ClustalX v2.1. For L1 ORF sequence data, the ClustalX alignment file was exported to IQTree webserver (Trifinopoulos et al., 2016) to determine the best model for a reliable estimate maximum likelihood phylogeny with an ultrafast bootstrap of 1000. This model was used to construct all maximum likelihood phylogenetic trees using MEGA X software. Additional sequence data for geographically distinct representatives of each HPV11 lineage, including lineages A, sub-lineages A1 to A4, and lineage B retrieved from GenBank were also included in downstream phylogenetic analysis.

Concatenated L1-URR and E5a/b-L1-URR sequences were generated with Geneious Prime software (https://www.geneious.com). To explore the phylogenetic relationship of the HPV isolates, maximum likelihood phylogenetic trees with a bootstrap of 1000 were constructed with the concatenated L1URR sequences and the concatenated E5a/b-L1-URR sequences. Phylogenetic trees were constructed with MEGA X software.

To explore whether the 208bp region in the E2 ORF sequenced in this study represents whole-genome variation, a Neighbor-Joining phylogenetic tree and a maximum likelihood phylogenetic tree with a bootstrap of 1000 were constructed with MEGA X software.

### 2.2.9 Ethical considerations

Ethical clearance was obtained from the Health Sciences Research Ethics Committee of the UFS. All isolates used in this study were collected by Professor Seedat (UFS-HSD2019/1109/2708) and informed consent was obtained to store the isolates and use them for related studies (ETOVS 194/2007 and ECUFS 6/2011). No additional isolates were collected in this study. Permission to perform the study was be obtained from the Head of the Division of Virology and the Head of the School of Pathology (National Health Laboratory Service).
2.3. Results

### 2.3.1 Patient data

During previous studies at the Division of Virology at the University of the Free State, Bloemfontein, laryngeal papilloma biopsies were collected from 2008 to 2018. DNA extracted from each biopsy was screened for HPV DNA and positive reactors were genotyped. Over this 10-year period, 94 HPV11 isolates from 70 patients diagnosed with RRP were identified. Eighty-one isolates from 60 patients were available for this study. Patient data is included in Appendix 1.

### 2.3.2 L1 ORF amplification and sequencing

PCR amplification of the L1 ORF was performed on all available isolates with primers HPV11-L1 and HPV11L1-R (Table 1). Eighty-one samples were purified, and separated with gel electrophoresis and 72 samples had a band at the predicted size of approximately 1770bp confirming the presence of HPV11 L1 DNA. The DNA concentration of the 72 samples ranged from $14.5 \mathrm{ng} / \mu \mathrm{l}$ to $202.1 \mathrm{ng} / \mu \mathrm{l}$. DNA concentrations of all isolates are included in Table 5.

Nucleotide sequences of each amplicon within the predicted size range was determined with Sanger sequencing. Nucleotide substitutions, insertions and deletions are summarised in Table 6. By comparing sequence data of isolates sequenced in the current study and the A1 prototype sequence L1 sequence, 86 nucleotide substitutions were observed in 11 genomic regions. This included a C6028T in 69 isolates (all except VBD34/08, VBD28/14 and VBD04/11), C6831T in one isolate (VBD34/08), A7126C in one isolate (VBD23/10), A7131C in one isolate (VBD34/08). Additionally, isolates VBD34/08 and VBD28/14 had A5929G, T6296A, T6298C, T6949A, G7045A C7197T, and T7237C nucleotide substitutions. The maximum number of nucleotide substitutions in a single isolate, VBD34/08, was nine which is $0,6 \%$ of the complete L1 ORF. Only one of the $72(1.4 \%)$ isolates, VBD04/11, was identical to the prototype.

Table 5: Quantitative estimation of DNA concentration revealed by Nanodrop spectrophotometer of human papillomavirus type 11 (HPV11) genomic regions including late region L1, upper regulatory region (URR), early region E5a/b and a segment of early region E2.
Isolates with insufficient DNA for quantitative estimation of DNA concentration are marked in dark grey.

| VBD | L1 (ng/ $\mu \mathrm{l}$ ) | URR (ng/ $/$ l) | E5a/b (ng/ul) | E2 segment (ng/ul) |
| :---: | :---: | :---: | :---: | :---: |
| 01/10 | 103,5 | 81,2 | 77,8 | 68,4 |
| 02/12 | 101,9 | 136,3 | 103,2 | 67,3 |
| 04/11 | 72,2 |  |  |  |
| 04/12 | 34 | 114 | 194 | 62,6 |
| 07/10 | 16,7 | 79,8 | 142,9 | 67 |
| 08/09 | 14,5 | 99,2 | 85,8 | 52,9 |
| 08/13 | 27,9 | 109,3 | 28,7 | 104 |
| 09/12 | 50,1 | 87,1 | 95,8 | 54,1 |
| 12/18 | 44,4 | 80,8 | 92,8 | 58,4 |
| 13/10 | 41,5 | 163,9 | 84,6 | 44 |
| 14/09 | 49,75 | 126,8 | 92,7 | 58 |
| 14/11 | 87,9 | 133,9 | 100,8 | 74,4 |
| 15/10 | 81,4 | 100,3 | 108,9 | 60,3 |
| 16/09 | 122,7 | 85,3 | 111,4 | 57,8 |
| 16/10 | 110,2 | 66,9 | 117,9 | 56,1 |
| 17/09 | 93,5 | 48,5 | 79,6 | 65,5 |
| 17/10 | 76,8 | 106,1 | 71,1 | 76,6 |
| 21/10 | 149,2 | 125,7 | 55,2 | 64,4 |
| 23/10 | 124,1 | 159,7 | 112,4 | 117,5 |
| 26/10 | 77,4 | 118 | 72 | 60,7 |
| 30/11 | 119,9 | 76,3 | 32,7 | 71,3 |
| 33/10 | 95,5 | 134,8 | 137,4 | 53,1 |
| 33/11 | 89,7 | 423,9 | 33,9 | 42,1 |
| 34/08 | 60,9 | 254 | 68,9 | 45,7 |
| 34/10 | 143,5 | 84,1 | 133,9 | 77,4 |
| 34/11 | 111,7 | 115,6 | 178,9 | 65,6 |
| 35/11 | 140,1 | 147,1 | 93 | 54,6 |
| 37/08 | 73,4 | 126 | 55,9 | 49 |
| 37/10 | 112,6 | 117,9 | 69,8 | 67,9 |
| 37/11 | 63,7 | 119,2 | 83,7 | 48,2 |
| 37/12 | 27,4 | 122,6 | 54,7 | 63,6 |
| 39/12 | 29,9 | 106,1 | 74,9 | 62,1 |
| 41/11 | 102,9 | 135,8 | 95,1 | 54,4 |
| 41/12 | 59,7 | 111,5 | 73,2 | 61,1 |
| 45/08 | 75,2 |  |  |  |
| 47/12 | 29 | 103,5 | 62,4 | 61,9 |
| 48/10 | 86,6 | 122,9 | 51,1 | 67,8 |
| 48/12 | 29,3 | 116,5 | 40,1 | 47,9 |
| 49/08 | 35 | 138,8 | 53,2 | 51,7 |
| 49/12 | 33,8 | 125,4 | 229,3 | 63,3 |
| 52/09 | 104,7 | 115,3 | 136,7 | 51,1 |
| 52/11 | 81,6 | 127,3 | 121,3 | 52,2 |
| 54/08 | 71,1 |  |  |  |
| 55/08 | 120,7 | 99,7 |  |  |
| 58/11 | 27,2 | 73,8 | 77 | 47,6 |
| 59/08 | 202,1 | 90,2 | 143,1 | 70,1 |
| 59/09 | 153 | 142,5 | 209,2 | 49,2 |
| 61/11 | 43 | 136,3 | 88,9 | 56,1 |
| 62/11 | 88 | 142,1 | 95,5 | 52,9 |
| 63/09 | 88,1 | 141,3 | 102,1 | 60,4 |
| 63/11 | 100,4 | 108,5 | 108,7 | 63,1 |
| 69/09 | 112,7 | 138,2 | 108,2 | 48,5 |
| 74/09 | 77,6 | 92,9 | 97,9 | 58,1 |
| 79/09 | 66,4 | 88 | 87,6 | 50,9 |
| 14/15 | 21,6 | 122,1 | 77,3 | 46 |
| 16/15 | 83,6 | 75,3 | 67 | 58,7 |
| 17/14 | 35,2 | 116,8 | 56,7 | 58,6 |
| 22/14 | 93,7 | 141,2 | 46,4 | 58,4 |
| 23/17 | 43,8 | 141,5 | 91,4 | 52,4 |
| 28/14 | 73,5 | 117,5 | 91,4 | 36,6 |
| 29/13 | 50,1 | 93,5 | 94,4 | 69,1 |
| 29/14 | 40 | 135 | 92,7 | 83,9 |
| 33/16 | 29,9 | 110,6 | 135,8 | 52,3 |
| 41/14 | 62 | 179,6 | 85,3 | 66 |
| 41/15 | 23,6 | 123,9 | 116,7 | 54,2 |
| 44/15 | 23,6 | 146,2 | 75,2 | 56,3 |
| 48/16 | 29,8 | 126,8 | 33,7 | 60,5 |
| 49/15 | 21,2 | 198,6 | 47,8 | 53,7 |
| 55/13 | 40,4 | 100,05 | 80,5 | 64,1 |
| 68/15 | 16 | 120,52 | 113,2 | 57,5 |
| 68/16 | 22,5 | 119,8 | 68,4 | 55,6 |
| 69/15 | 21,7 |  |  |  |

Table 6: Sequence comparison of nucleotide variances among 72 human papillomavirus type 11 (HPV11) isolates L1 region against the HPV11 prototype with GenBank accession number M14119.1.
Residues that differed to those of HPV11 L1 prototype are indicated by a white colour difference and the substituted base.


### 2.3.3 URR amplification and sequencing

Following the amplification of the L1 ORF, four isolates had insufficient DNA and were excluded from DNA amplification and nucleotide determination during downstream analysis. These included VBD04/11, VBD69/15, VBD45/08 and VBD54/08.

PCR amplification of the URR was performed on the available isolates with primers summarised in Table 1. The remaining 68 samples were purified, and $5 \mu$ aliquots of the purified HPV11 DNA were separated with gel electrophoresis. The gel was visualised and 68 samples had a band at the predicted size confirming the presence of HPV11 URR DNA. The DNA concentration of the 68 samples ranged from $48.5 \mathrm{ng} / \mu \mathrm{l}-423.9 \mathrm{ng} / \mu$. DNA concentrations of all isolates are included in Table 5.

Nucleotide sequences of each amplicon within the predicted size range were determined with Sanger sequencing using the same primers as for the initial PCR (Table 1) plus an additional primer HPV11-LCR-FF. By comparing sequence data of isolates sequenced in the current study and the A1 prototype sequence URR sequence, 237 nucleotide substitutions could be observed in 23 genomic regions. Nucleotide substitutions, insertions, and deletions in the URR are summarised in Table 7. All isolates had a T7547C nucleotide substitution. Isolates VBD28/14 and VBD34/08 had A7289G, A7302C, G7319T, G7329C, G7349C, T7359C, T7411G, G7519T, C7536T, A7570C, A7591G, A7626C, T7646A, T7775G, G7780T, C7880T, C7928T, G29T, and C50T nucleotide substitutions. Other nucleotide substitutions included C7333A in 10 isolates, A7413C in 55 isolates, C7479T in 66 isolates.

The maximum number of nucleotide substitutions per isolate was 20 in VBD34/08 and VBD28/14. Five nucleotide substitutions were detected in isolates VBD33/16, VBD68/15, VBD41/15, VBD41/14, VBD35/11, VBD59/09, VBD63/09, VBD16/09, VBD52/09, VBD37/10. A total of 118 nucleotide insertions were present among 58 isolates. The majority of isolates had a $\mathrm{C}(50 / 68,73.5 \%)$ or $\mathrm{CC}(7 / 68$, $10.3 \%$ ) insertion following genomic position 7575 (All except VBD44/15, VBD29/14, VBD55/13, VBD47/12, VBD39/12, VBD37/11, VBD34/11, VBD23,10, VBD74/09, VBD15/10, and VBD34/08). Only isolates with a C or CC insertion following genomic position 7575 had no nucleotide substitution at genomic position 7413.

The maximum number of nucleotide insertions in an isolate was 28 in VBD28/14 in five genomic regions, followed by 27 in VBD34/08 in four genomic regions. Insertions in VBD28/14 included a GCACGC insertion following genomic position 7527, a C insertion following genomic position 7575, a GGCGCCA insertion following genomic position 7715, a TGGGTTG insertion following genomic position 7744, and a TTATCTC insertion following genomic position 7746. Insertions in VBD34/08 included a GCACGC insertion following genomic position 7527, a GGCGCCA insertion following

Table 7: Sequence comparison of nucleotide variances among 68 human papillomavirus type 11 (HPV11) isolates upper regulatory region (URR) against the HPV11 prototype with GenBank accession number M14119.1.
For the URR, a corrected sequence with a $2 b p$ insertion at nt 7717-7718 was used when determining genomic variants and genomic positions (Maver et al., 2011). Residues that differed to those of HPV11 URR prototype are indicated by colour differences. White depicts nucleotide substitutions, yellow depicts nucleotide insertions and green depicts nucleotide deletions. Samples excluded from analysis are indicated in dark grey.

genomic position 7715, a TGGGTTG insertion following genomic position 7744, and a TTATCTC insertion following genomic position 7746.

There were 90 nucleotide deletions observed in the 68 isolates; 66 of the $68(97 \%)$ isolates had a single nucleotide deletion at genomic position 7506 and isolates VBD28/14 and VBD34/08 each had 12 nucleotide deletions at genomic positions 7682, 7684-7687, 7692-7697 and 7699. No isolates were identical to the prototype (GenBank accession number M14119.1).

### 2.3.1 E5a/b ORF amplification and sequencing

PCR amplification of the E5a/b ORF were performed on all available isolates with primers summarised in Table 1. One isolate, VBD55/08, had insufficient DNA for downstream analysis.

Samples were purified, and $5 \mu \mathrm{l}$ aliquots of the purified HPV11 DNA were separated with gel electrophoresis and visualised. Sixty-seven isolates had a band at the predicted size confirming the presence of HPV11 E5a/b DNA. The DNA concentration of the isolates ranged from $28.7 \mathrm{ng} / \mu \mathrm{l}-229.3$ $\mathrm{ng} / \mu \mathrm{l}$. All isolates DNA concentration are included in Table 5.

Nucleotide sequences of each amplicon within the predicted size range were determined with Sanger sequencing using the same primers as for the initial PCR (Table 1). By comparing sequence data of isolates sequenced in the current study and the A1 prototype sequence $\mathrm{E} 5 \mathrm{a} / \mathrm{b}$ sequence, 166 nucleotide substitutions could be observed in 15 genomic regions (Table 8). These included A3891C in one isolate (VBD23/10), A3952T and G3991C in all 67 isolates, A4142C in 11 isolates, G4325T in one isolate (VBD48/10), and A4344G in one isolate (VBD68/16). Additionally, isolates VBD34/08 and VBD28/14 had C3902G, A3978G, A3993T, T4048C, C4166G, C4227T, A4274C, C4312A, and C4333T nucleotide substitutions.

The maximum number of nucleotide substitutions was 11 in isolates VBD28/14 and VBD34/08, which is $2.2 \%$ of the complete E5a/b ORF, followed by four nucleotide substitutions in VBD23/10, which is $0.8 \%$ of the complete E5a/b ORF. All isolates had nucleotide substitutions at genomic position 3952 and 3991. No nucleotide insertions or nucleotide deletions were observed compared to GenBank accession number M14119.

Table 8: Sequence comparison of nucleotide variances among 67 human papillomavirus type 11 (HPV11) isolates E5a/b region against the HPV11 prototype with GenBank accession number M14119.1.
Residues that differed to those of HPV11 E5a/b prototype are indicated by a white colour difference and the substituted base.
Samples excluded from analysis are indicated in dark grey.


### 2.3.2 E2 ORF amplification and sequencing

PCR amplification of the E2 ORF segment was performed on all available isolates with designed primers (Table 1). The 67 available samples were purified, and a $5 \mu 1$ aliquot of the purified HPV11 DNA was separated with gel electrophoresis. All the isolates had a band at the predicted size confirming the presence of target DNA. The DNA concentration ranged from $36.6 \mathrm{ng} / \mu \mathrm{l}-117.5 \mathrm{ng} / \mu \mathrm{l}$. DNA concentrations of all isolates are included in Table 5.

Nucleotide sequences of each amplicon were determined with Sanger sequencing using the same primers as for the initial PCR (Table 1). Among the 67 isolates, 144 nucleotide substitutions were observed by comparing sequence data of isolates sequenced in the current study and the E2 segment of the A1 prototype. Nucleotide substitutions are summarised in Table 9. Nucleotide substitutions were observed in 11 genomic regions. These included A3633C in 65 isolates (all except VBD34/08 and VBD28/14), A3652G in all 67 isolates, T3672C in one isolate (VBD62/11), C3705A in one isolate (VBD61/11), A3738T in one isolate (VBD14/11), G3758C in one isolate (VBD29/14), A3779C is one isolate (VBD74/09), and A3827G in one isolate (VBD34/08). Isolates VBD34/08 and VBD28/14 had G3638C, C3788T, A3821G nucleotide substitutions.

The maximum number of nucleotide substitutions in the 208bp gene segment of the E 2 was five (5/208, $2.4 \%$ ) nucleotide substitutions in VBD34/08 and four (4/208, 1.9\%) nucleotide substitutions in VBD28/14. No nucleotide insertions or nucleotide deletions were observed. No isolates were identical to the prototype (GenBank accession number M14119.1).

### 2.3.3 Analysis of amino acid variations

## Analysis of LI ORF sequences

The L1 gene contains 501 amino acids. By comparing the amino acid of isolates sequenced in the current study and the A1 prototype, six nucleotide substitutions were identified as synonymous, and five as non-synonymous. Non-synonymous nucleotide substitutions resulted in four amino acid substitutions.

Nucleotide substitutions changed the amino acid sequence of two isolates, namely VBD28/14 and VBD34/08. Both isolates contained T6296A and T6298C nucleotide substitutions causing a serine (S) to a threonine (T) amino acid substitution and a C7197T nucleotide substitution causing an alanine (A) to a valine amino acid substitution and only VBD34/08 contained a C6831T nucleotide substitution causing a S to a leucine ( L ) amino acid substitution and a A7131C nucleotide substitution causing a glutamic acid (E) to an A.

Table 9: Sequence comparison of nucleotide variances among 67 human papillomavirus type 11 (HPV11) isolates E2 segment against the HPV11 prototype with GenBank accession number M14119.1.
Residues that differed to those of HPV11 E2 segment prototype are indicated by a white colour difference and the substituted base.

|  | E2 Genomic region |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
|  | 6 | 6 | 6 | 6 | 7 | 7 | 7 | 7 | 7 | 8 | 8 |
|  | 3 | 3 | 5 | 7 | 0 | 3 | 5 | 7 | 8 | 2 | 2 |
|  | 3 | 8 | 2 | 2 | 5 | 8 | 8 | 9 | 8 | 1 | 7 |
| Base | A | G | A | T | C | A | G | A | C | A | A |
| Isolate VBD <br> Number |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| 12/18 | C |  | G |  |  |  |  |  |  |  |  |
| 23/17 | C |  | G |  |  |  |  |  |  |  |  |
| 68/16 | C |  | G |  |  |  |  |  |  |  |  |
| 48/16 | C |  | G |  |  |  |  |  |  |  |  |
| 33/16 | C |  | G |  |  |  |  |  |  |  |  |
| 68/15 | C |  | G |  |  |  |  |  |  |  |  |
| 49/15 | C |  | G |  |  |  |  |  |  |  |  |
| 41/15 | C |  | G |  |  |  |  |  |  |  |  |
| 44/15 | C |  | G |  |  |  |  |  |  |  |  |
| 16/15 | C |  | G |  |  |  |  |  |  |  |  |
| 14/15 | C |  | G |  |  |  |  |  |  |  |  |
| 41/14 | C |  | G |  |  |  |  |  |  |  |  |
| 28/14 |  | C | G |  |  |  |  |  | T | G |  |
| 17/14 | C |  | G |  |  |  |  |  |  |  |  |
| 22/14 | C |  | G |  |  |  |  |  |  |  |  |
| 29/14 | C |  | G |  |  |  | C |  |  |  |  |
| 55/13 | C |  | G |  |  |  |  |  |  |  |  |
| 29/13 | C |  | G |  |  |  |  |  |  |  |  |
| 08/13 | C |  | G |  |  |  |  |  |  |  |  |
| 48/12 | C |  | G |  |  |  |  |  |  |  |  |
| 47/12 | C |  | G |  |  |  |  |  |  |  |  |
| 39/12 | C |  | G |  |  |  |  |  |  |  |  |
| 37/12 | C |  | G |  |  |  |  |  |  |  |  |
| 09/12 | C |  | G |  |  |  |  |  |  |  |  |
| 61/11 | C |  | G |  | A |  |  |  |  |  |  |
| 62/11 | C |  | G | C |  |  |  |  |  |  |  |
| 58/11 | C |  | G |  |  |  |  |  |  |  |  |
| 52/11 | C |  | G |  |  |  |  |  |  |  |  |
| 41/11 | C |  | G |  |  |  |  |  |  |  |  |
| 37/11 | C |  | G |  |  |  |  |  |  |  |  |
| 35/11 | C |  | G |  |  |  |  |  |  |  |  |
| 49/12 | C |  | G |  |  |  |  |  |  |  |  |
| 30/11 | C |  | G |  |  |  |  |  |  |  |  |
| 34/11 | C |  | G |  |  |  |  |  |  |  |  |
| 14/11 | C |  | G |  |  | T |  |  |  |  |  |
| 04/11 |  |  |  |  |  |  |  |  |  |  |  |
| 48/10 | C |  | G |  |  |  |  |  |  |  |  |
| 33/10 | C |  | G |  |  |  |  |  |  |  |  |
| 63/11 | C |  | G |  |  |  |  |  |  |  |  |
| 02/12 | C |  | G |  |  |  |  |  |  |  |  |
| 04/12 | C |  | G |  |  |  |  |  |  |  |  |
| 26/10 | C |  | G |  |  |  |  |  |  |  |  |
| 21/10 | C |  | G |  |  |  |  |  |  |  |  |
| 23/10 | C |  | G |  |  |  |  |  |  |  |  |
| 01/10 | C |  | G |  |  |  |  |  |  |  |  |
| 07/10 | C |  | G |  |  |  |  |  |  |  |  |
| 17/10 | C |  | G |  |  |  |  |  |  |  |  |
| 34/10 | C |  | G |  |  |  |  |  |  |  |  |
| 79/09 | C |  | G |  |  |  |  |  |  |  |  |
| 74/09 | C |  | G |  |  |  |  | C |  |  |  |
| 15/10 | C |  | G |  |  |  |  |  |  |  |  |
| 59/09 | C |  | G |  |  |  |  |  |  |  |  |
| 63/09 | C |  | G |  |  |  |  |  |  |  |  |
| 69/15 |  |  |  |  |  |  |  |  |  |  |  |
| 55/08 |  |  |  |  |  |  |  |  |  |  |  |
| 34/08 |  | C | G |  |  |  |  |  | T | G | G |
| 37/08 | C |  | G |  |  |  |  |  |  |  |  |
| 08/09 | C |  | G |  |  |  |  |  |  |  |  |
| 14/09 | C |  | G |  |  |  |  |  |  |  |  |
| 41/12 | C |  | G |  |  |  |  |  |  |  |  |
| 17/09 | C |  | G |  |  |  |  |  |  |  |  |
| 49/08 | C |  | G |  |  |  |  |  |  |  |  |
| 59/08 | C |  | G |  |  |  |  |  |  |  |  |
| 69/09 | C |  | G |  |  |  |  |  |  |  |  |
| 16/10 | C |  | G |  |  |  |  |  |  |  |  |
| 16/09 | C |  | G |  |  |  |  |  |  |  |  |
| 52/09 | C |  | G |  |  |  |  |  |  |  |  |
| 37/10 | C |  | G |  |  |  |  |  |  |  |  |
| 33/11 | C |  | G |  |  |  |  |  |  |  |  |
| 45/08 |  |  |  |  |  |  |  |  |  |  |  |
| 13/10 | C |  | G |  |  |  |  |  |  |  |  |
| 54/08 |  |  |  |  |  |  |  |  |  |  |  |

## Analysis of E5a/b ORF sequences

The E5a gene (nt 3871-4146) contains 91 amino acids and the E5b gene (nt 4146-4370) contains 73 amino acids. Three nucleotide substitutions were identified as synonymous and 12 substitutions were identified as non-synonymous by comparing the amino acid of isolates sequenced in the current study and the A1 prototype. Eleven amino acid substitutions were observed. Non-synonymous substitutions included A3891C nucleotide substitution causing a glutamine (Q) to isoleucine (I) amino acid substitution in one isolate (VBD23/10), A4142C nucleotide substitution causing Q to proline $(\mathrm{P})$ amino acid substitution in 11 isolates (VBD44/15, VBD29/14, VBD55/13, VBD47/12, VBD39/12, VBD41/11, VBD37/11, VBD34/11, VBD23/10, VBD74/09 and VBD15/10), and A4344G nucleotide substitution causing a N to aspartic acid ( D ) amino acid substitution in one isolate (VBD68/16). All isolates contained a A3952T and G3991C nucleotide substitutions causing an I to phenylalanine (F) and a valine to $L$ amino acid substitution, respectively.

Isolates VBD28/14 and VBD34/08 had the most non-synonymous substitutions including C3902G nucleotide substitution causing an A to glycine (G) amino acid substitution, A3952T nucleotide substitution causing an I to F amino acid substitution, G3991C and A3993T nucleotide substitution causing valine to L amino acid substitution, C 4166 G nucleotide substitution causing a histidine to Q amino acid substitution, C 4227 T nucleotide substitution causing a L to F amino acid substitution, A4274C nucleotide substitution causing a lysine $(\mathrm{K})$ to asparagine ( N ) amino acid substitution, C 4312 A nucleotide substitution causing a T to N amino acid substitution, and C 4333 T nucleotide substitution causing a S to L amino acid substitution.

### 2.3.4 Phylogenetic analysis

Four phylogenetic trees were constructed to explore the phylogenetic relationship of the HPV11 isolates sequenced in this study. Neighbor-Joining phylogenetic trees and maximum likelihood trees were constructed with sequences obtained in this study and 28 sequences retrieved from GenBank of geographically distinct HPV isolates belonging to different (sub)lineages.

Isolates with GenBank accession numbers M14119.1, LN833187.1, LN833185.1, LN833184.1, LN833183.1, LN833169.1, LN833165.1, LN833161.1, KU298879.1, JQ773412.1, JQ773411.1, JQ773409.1, JQ773408.1, JN644141.1, HE611263.1, HE574702.1, FR872717.1, FN907963.1, FN907962.1, FN870021.1, EU918768.1, MN788368.1, MK463921.1, MK463916.1, MK463914.1, MK313767.1, MK313765.1, and MK313763.1 were included. Lineage and sub-lineage representatives are summarised in Table 10.

Table 10: GenBank accession numbers of human papillomavirus type 11 (HPV11) isolates and the (sub)lineage representation.

| GenBank Accession number | Lineage and sub-lineage representatives |  |
| :---: | :---: | :---: |
| M14119.1 | Sub-lineage A1 |  |
| LN833161.1 | Sub-lineage A2 |  |
| LN833169.1 | Sub-lineage A3 |  |
| LN833187.1 | Sub-lineage A4 |  |
| LN833183.1 |  | Lineage B |

## Maximum Likelihood method

The evolutionary history was deduced using the maximum likelihood method. The Tamura-Nei model was determined as the best model for a reliable estimate maximum likelihood phylogeny with an ultrafast bootstrap of 1000 . The trees with the highest log-likelihood for concatenated L1-URR ($10510,15)$ and E5a/b-L1-URR $(-11642,10)$ are depicted in Figure 4 and Figure 5, respectively.

Analysis of concatenated L1-URR involved 96 nucleotide sequences, including 68 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank. Results revealed that analysis of concatenated L1-URR resolved into two lineages, namely the lineage A and lineage B. The majority of isolates $(49 / 68,72.0 \%)$ clustered together with a $94 \%$ certainty in the bootstrap test.

Analysis of concatenated E5a/b-L1-URR involved 95 nucleotide sequences. Sixty-seven sequences obtained in this study, and 28 sequences retrieved from GenBank from geographically distinct isolates. HPV11 concatenated E5a/b-L1-URR resolved into two lineages, namely lineage A and lineage B. The majority of the isolates sequenced in this study (48/67, 71.6\%) clustered with the A2 prototype LN833161.1 and revealed a $97 \%$ certainty in the bootstrap test. Seven isolates from this study, clustered with one isolate (retrieved from GenBank) previously identified as an A2 sub-lineage isolate with an $82 \%$ certainty in the bootstrap test (Jelen et al., 2016).

Three isolates previously identified as A1 sub-lineage isolates (EU918768.1, FN870021.1 and FN907963.1), did not closely cluster together. No isolates sequenced in this study clustered with the A3 and A4 prototype strains.

Figure 4: Phylogenetic tree of human papillomavirus type 11 (HPV11) based on alignment of 96 nucleotide sequences of HPV11 concatenated L1-URR. This analysis involves 68 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.
M14119.1 (blue) was used as a representative of sublineage A1, LN833161.1 (green) as a representative of sub-lineage A2, LN833169.1 (yellow) as a representative of sub-lineage A3, LN833187.1 (red) as a representative of sub-lineage A4. Lastly, LN833183.1 was used as a representative of lineage $B$. The maximum likelihood tree was constructed using MEGA X using the Maximum Likelihood method and TamuraNei model.

Ten isolates from this study (VBD15/10, VBD29/14, VBD34/11, VBD37/11, VBD39/12, VBD44/15, VBD47/12, VBD55/13, VBD23/10 and VBD74/09) clustered with two (MK463921.1 and MK463916.1) sequences retrieved from GenBank with $99 \%$ and $100 \%$ certainty in the bootstrap test in HPV11 concatenated L1-URR and E5a/b-L1-URR data sets, respectively. The sub-lineage classification of these GenBank sequences are unknown.

Seven isolates from the current study (VBD01/10, VBD07/10, VBD17/10, VBD17/14, VBD22/14, VBD33/11, and VBD34/10) clustered with isolates FN907963.1 and FN907962.1 with a $98 \%$ and $99 \%$ certainty in the bootstrap test in HPV11 concatenated L1-URR and E5a/b-L1-URR data sets, respectively. However, FN907963.1 has been previously identified as sub-lineage A1 and FN907962.1 as sub-lineage A2. One confirmed A1 sub-lineage isolate (EU918768.1) retrieved from GenBank shared a node with confirmed A2 sub-lineage isolates with a $95 \%$ certainty in the bootstrap test. In both concatenated L1-URR and E5a/b-L1-URR sequences, three lineage B isolates, one from GenBank (LN833183.1) and two from the current study (VBD34/08 and VBD28/14), clustered together in the bootstrap test with $100 \%$ certainty with one isolate being identical to the previously identified lineage $B$ isolate.

## Pairwise analysis of percentage divergence of nucleotides

Analyses were performed with the use of the Maximum Composite Likelihood model and included 95 nucleotide sequences and used the concatenated E5a/b-L1-URR data set. This includes 67 sequences obtained in this study, and 28 sequences retrieved from GenBank. All numbers were limited to three decimals. The pairwise analysis of percentage divergence of nucleotides using the human papillomavirus type 11 concatenated E5a/b-L1-URR data set is included in Appendix 10.

Divergence from the A1 prototype M14119.1 ranged from minimum 0,087 (EU918768.1 and FN870021.1) to a maximum of 0,154 (VBD34/08, FN907962.1 and LN833183.1) with an average divergence of 0,119 . Divergence from the A2 prototype LN833161.1 ranged from 0,001 (VBD02/12, VBD04/12, VBD08/09, VBD08/13, VBD09/12, VBD12/18, VBD13/10, VBD14/09, VBD14/11, VBD14/15, VBD16/10, VBD16/15, VBD17/09, VBD21/10, VBD23/17, VBD26/10, VBD29/13, VBD30/11, VBD33/10, VBD37/08, VBD37/12, VBD41/11, VBD41/12, VBD48/12, VBD48/16, VBD49/08, VBD49/12, VBD49/15, VBD52/11, VBD58/11, VBD59/08, VBD61/11, VBD62/11, VBD63/11, VBD69/09, HE611263.1, JN644141.1, JQ773409.1, MK313763.1, MN788368.1, VBD79/09) to $0,166(V B D 34-08$ and LN833183.1) with an average divergence of 0,037 . Divergence from the A3 prototype LN833169.1 ranged from 0,000 (LN833165.1) to 0,158 (LN833183.1) with an average divergence of 0,058 . Divergence from the A4 prototype LN833187.1 ranged from 0,005 (LN833165.1 and LN833169.1) to 0,159 (VBD34/08 and LN833183.1) with an average divergence of 0,060. Divergence from the lineage B prototype LN833183.1 ranged from 0,000 (VBD34/08) to 0,166


Figure 5: Phylogenetic tree of human papillomavirus type 11 (HPV11) based on alignment of 95 nucleotide sequences of HPV11 concatenated E5a/b-L1-URR. This analysis involves 67 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.
M14119.1 (blue) was used as a representative of sublineage A1, LN833161.1 (green) as a representative of sub-lineage A2, LN833169.1 (yellow) as a representative of sub-lineage A3, LN833187.1 (red) as a
Lineage B representative of sub-lineage A4. Lastly, LN833183.1 (orange) was used as a representative of lineage B. The maximum likelihood tree was constructed using MEGA $X$ using the Maximum Likelihood method and TamuraNei model.
(VBD16/09, VBD33/16, VBD35/11, VBD37/10, VBD41/14, VBD41/15, VBD48/10, VBD52/09, VBD59/09, VBD63/09, VBD68/15, VBD68/16, FN870021.1, FR872717.1, JQ773408.1, JQ773409.1, JQ773411.1, JQ773412.1, KU298879.1, LN833161.1, LN833165.1, LN833169.1, LN833183.1, LN833184.1, LN833185.1, M14119.1, MK313763.1, MK313765.1, MK313767.1, MK463914.1) with an average divergence of 0,161 . VBD28/14 had a divergence of 0,120 from LN833183.1 and VBD34/08 (Appendix 10).

Analysis of E2 segment as representative of whole genome variation

The evolutionary history was inferred with the Neighbor-Joining method and the Maximum Composite Likelihood method and involved 95 nucleotide sequences. There were a total of 208 nucleotide positions in the final dataset. The phylogenetic trees are shown in Figure 6 and Figure 7.

HPV11 E2 segment sequences clearly resolved into two lineages, namely the lineage A and lineage B. Sixty-five of the 67 isolates $(97.0 \%)$ sequenced clustered together with the A2 prototype LN833161.1 and revealed a $52 \%$ and $56 \%$ certainty in the bootstrap test the Neighbor-Joining and maximum likelihood phylogenetic trees, respectively. Lineage B isolates clustered together with a $95 \%$ certainty in the bootstrap test in both the Neighbor-Joining and maximum likelihood phylogenetic trees.

Four isolates clustered with the A1 prototype M14119.1 with $54 \%$ and $59 \%$ confidence according to the Neighbor-Joining and maximum likelihood methods, respectively. Analysis of estimates of evolutionary divergence between sequences revealed the minimum divergence from the A1 prototype M14119.1 was 0,000 (EU918768.1, FN870021.1, and FN907963.1), the maximum was 0,024 (VBD34/08 and LN833183.1) (Appendix 11).

Divergence from A2 prototype LN833161.1 ranged from 0,000 (VBD01/10, VBD02/12, VBD04/12, VBD07/10, VBD08/09, VBD08/13, VBD09/12, VBD12/18, VBD13/10, VBD14/09, VBD14/15, VBD15/10, VBD16/09, VBD16/10, VBD16/15, VBD17/09, VBD17/10, VBD17/14, VBD21/10, VBD22/14, VBD23/10, VBD23/17, VBD26/10, VBD29/13, VBD30/11, VBD33/10, VBD33/11, VBD33/16, VBD34/10, VBD34/11, VBD35/11, VBD37/08, VBD37/10, VBD37/11, VBD37/12, VBD39/12, VBD41/11, VBD41/12, VBD41/14, VBD41/15, VBD44/15, VBD47/12, VBD48/10, VBD48/12, VBD48/16, VBD49/08, VBD49/12, VBD49/15, VBD52/09, VBD52/11, VBD55/13, VBD58/11, VBD59/08, 59/09, VBD63/09, VBD63/11, VBD68/15, VBD68/16, VBD69/09, VBD69/15, VBD79/09, HE574702.1, HE611263.1, JN644141.1, JQ773409.1, KU298879.1, LN833184.1, LN833185.1, MK313763.1, MK313765.1, MK313767.1, MN788368.1) to 0,020 (VBD34/08 and LN833183.1) (Appendix 11).

No isolates sequenced in the current study clustered with the A3 or A4 prototype sequences. Divergence from the A3 prototype LN833169.1 ranged from 0,000 (LN833165.1) to 0,024 (VBD34/08 and

LN833183.1) (Appendix 11). The A4 prototype LN833187.1 contains a wobble nucleotide in the 208bp region of the E2 ORF. One nucleotide was substituted with the degenerative nucleotide ' K ', which can be interpreted as either a G or a T . Therefore, the trees were constructed using both a G and a T and allocated LN833187.1(1) and LN833187.1 (2), respectively (Appendix 4).

Divergence from LN833187.1_(1) ranged from 0,005 (VBD01/10, VBD02/12, VBD04/12, VBD07/10, VBD08/09, VBD08/13, VBD09/12, VBD12/18, VBD13/10, VBD14/09, VBD14/15, VBD15/10, VBD16/09, VBD16/10, VBD16/15, VBD17/09, VBD17/10, VBD17/14, VBD21/10, VBD22/14, VBD23/10, VBD23/17, VBD26/10, VBD29/13, VBD30/11, VBD33/10, VBD33/11, VBD33/16, VBD34/10, VBD34/11, VBD35/11, VBD37/08, VBD37/10, VBD37/11, VBD37/12, VBD39/12, VBD41/11, VBD41/12, VBD41/14, VBD41/15, VBD44/15, VBD47/12, VBD48/10, VBD48/12, VBD48/16, VBD49/08, VBD49/12, VBD49/15, VBD52/09, VBD52/11, VBD55/13, VBD/11, VBD59/08, VBD59/09, VBD63/09, VBD63/11, VBD68/15, VBD68/16, VBD69/09, VBD69/15, VBD79/09, EU918768.1, FN870021.1, FN907963.1, HE574702.1, HE611263.1, JN644141.1, JQ773409.1 KU298879.1, LN833161.1, LN833165.1, LN833169.1, LN833184.1, LN833185.1, LN833187.1_(2), M14119.1, MK313763.1, MK313765.1, MK313767.1, MN788368.1) to 0,020 (VBD34/08 and LN833183.1) (Appendix 11).

Divergence from LN833187.1_(2) ranged from 0,005 (LN833187.1_(1)) to 0,024 (VBD34/08 and LN833183.1) with an average divergence of 0,011 . Divergence from LN833183.1 ranged from 0,000 (VBD34/08) to 0,034 (FR872717.1, JQ773408.1, JQ773411.1, JQ773412.1, and MK463914.1) (Appendix 11).

Three lineage B isolates (VBD34/08, VBD28/14 and, LN833183.1) clustered together in the bootstrap test with $95 \%$ certainty according to the bootstrap test in both the Neighbor-Joining and maximum likelihood phylogenetic trees. Divergence from ranged from 0,000 (VBD34/08) to 0,034 (FR872717.1, JQ773408.1, JQ773411.1, JQ773412.1 (Appendix 11).


Figure 6: Phylogenetic tree of human papillomavirus type 11 (HPV11) based on 95 nucleotide sequence alignments of HPV11 E2 segment. This analysis involves 67 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.

M14119.1 (blue) was used as a representative of sub-lineage A1, LN833161.1 (green) as a representative of sub-lineage A2, LN833169.1 (yellow) as a representative of sub-lineage A3, LN833187.1 (red) as a representative of sub-lineage A4. Lastly, LN833183.1 (orange) was used as a representative of lineage B. A single, double peak was denoted as a wobble nucleotide in the LN833187.1 genome sequence submitted to GenBank, and therefore for readability, the trees were constructed using both a G and a $T$ and allocated $L N 833187.1(1)$ and $L N 833187.1$ (2), respectively. The maximum likelihood tree was constructed using MEGA X using the Maximum Likelihood method and Tamura-Nei model.


Figure 7: Phylogenetic tree of human papillomavirus type 11 (HPV11) based on 95 nucleotide sequence alignments of HPV11 E2 segment. This analysis involves 67 isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank.

M14119.1 (blue) was used as a representative of sub-lineage A1, $L N 833161.1$ (green) as a representative of sub-lineage A2, LN833169 (yellow) as a representative of sub-lineage A3, LN833187.1 (red) as a representative of sub-lineage A4. Lastly, LN833183.1 (orange) was used as a representative of lineage B. A double peak was denoted as a wobble nucleotide in the LN833187.1 genome sequence submitted to GenBank, hence, the tree was constructed using both a $G$ and a $T$ and allocated LN833187.1(1) and LN833187.1 (2), respectively. The Neighbor-Joining tree was constructed using MEGA X.

### 2.4. Discussion

During the course of this study, genome sequencing was used to identify unique HPV11 variants circulating in patients diagnosed with RRP treated at the Universitas Academic Hospital. HPV11 has a prevalence of $1 \%$ to $4 \%$ in cervical samples in various South African regions (Giuliano et al., 2015; Mbulawa et al., 2018). Several studies have reported HPV prevalence from different parts of South Africa (Giuliano et al., 2015; Mbulawa et al., 2018). However, very few studies have focused on the prevalence of HPV11 genetic variants in South Africa. Therefore, in the current study, we aimed to investigated HPV11 variants circulating in the study population by means of genomic sequencing of the L1, E5a/b and URR genes, which have been postulated to produce phylogenetic trees able to distinguish between lineages and sub-lineages (Jelen et al., 2016). In addition, the hypothesis that a 208bp segment of the E2 genome generates tree topology which represents whole-genome variation was also challenged (Jelen et al., 2016). HPV11 variant investigation is essential, as it is unknown if genome variations may result in divergent infectivity rates, affects vaccine effectiveness, and disease prognosis.

A previous study aiming to identify HPV6 variants in patients diagnosed with RRP in South Africa identified three novel HPV6 variants (Combrinck et al., 2012). Subsequent studies focusing on the molecular characterisation of HPV6 suggested that HPV6 sub-lineage B1 is associated with an increased risk of developing genital warts (Flores-Díaz et al., 2017). HPV11 variants circulating in South Africa are currently largely unknown and the common occurrence of HPV6 variants in patients diagnosed with RRP in South Africa, suggest that similar patterns may be observed with HPV11. Therefore, further investigative studies with regards to HPV 11 are warranted.

Between 2008 and 2018, 94 HPV11 isolates were identified in patients diagnosed with RRP at the Universitas Academic Hospital in the Free State province of South Africa. Previously, a novel lineage $B$ isolate was identified, necessitating further investigation (Makatsa, 2012). However, this isolate was not available for characterisation in the current study. In this study, the concatenated L1-URR and E5a/b-L1-URR genes were characterised genetically to discriminate between different (sub)lineages. Despite the proof-reading abilities of host cell polymerases, many SNPs were identified in isolates characterised in this study which likely reflect a difference in the HPV11 genome.

Within the L1 ORF, which is generally considered a conserved region of the HPV genome, sequence variation was low, likely due to the slow genomic evolutionary rate, which is estimated to be $10^{-8}$ nucleotide substitutions per site per year (Chen et al., 2009, 2011). The highest percentage of divergence from the A1 prototype M14119.1 was $0.6 \%$, implying a variant sub-lineage as this region is often used to represent whole-genome variation (Burk et al., 2011). Ninety-five per cent (69/72) of sequences incorporating a C6028T substitution and only one isolate was identical to the A1 prototype M14119.1.

However, the L1 gene is generally only used to identify the type of HPV, and additional sequencing is required to identify variant lineages and sub-lineages reliably (Burk et al., 2011).

Although nucleic acid changes may imply a variant lineage or sub-lineage, nucleic acid variances do not necessarily impact the function of the expressed protein. DNA alterations in an ORF may lead to alterations in the gene product if non-synonymous substitutions are present (Lebeuf-Taylor et al., 2019). Analysis of genome variances in the L1 ORF identified four non-synonymous substitutions in two isolates, namely VBD28/14 and VBD34/08. The L1 major capsid protein is highly immunogenic and has formed the basis of successful vaccines (Chabeda et al., 2018; Dadar et al., 2018). The L1 also mediates vital functions for virus survival, such as encapsulation of the papillomavirus genome, interaction with the host cell for infectious entry, and releasing the viral DNA into a new host cell (Barra et al., 2019; Benedict \& Derkay, 2021). Consequently, amino acid changes may impact virus survival and the effectiveness of HPV11 targeted vaccines.

Typically, evaluation of vaccine antibody responses is studied using serological assays. Sites of variation for the L1 region of HPV11 isolates analysed in this study were confined to a limited number of internal residues implying that the influence of these variances on recognition by L1-specific antibodies is likely to be insignificant, as indicated in a study on HPV16 (Pastrana et al., 2004). The consequences of these polymorphisms in the novel lineage B and sub-lineage A3 and A4 and antibody recognition following natural infection or vaccination by the present-day vaccines requires additional investigation.

For HPV11 sub-lineage classification, sequencing of additional gene segments is required. Genomic sequencing of the $\mathrm{E} 5 \mathrm{a} / \mathrm{b}$ revealed a $2.2 \%$ divergence from the A1 prototype M14119.1 in VBD28/14 and VBD34/08, and a $0.8 \%$ divergence from the prototype in VBD23/10. Most of the nucleotide substitutions were non-synonymous and resulted in amino acid alterations, which may interfere with the function of this protein (Lebeuf-Taylor et al., 2019). Proteins expressed by the E5a/b are each made up of 83 amino acids and are involved in many processes, including immune evasion, and regulating apoptosis. The E5a/b proteins also play a role in cell cycle pathways and influence cellular gene expression (Chagas et al., 2011; Zhang et al., 2018).

Few studies have reported on gene mutations of E5 in HPV11, but studies on various HPV types have reported that non-synonymous mutations might lead to changes in polarity, hydropathic potential, and the amino acid side chain, which may alter protein folding (Halavaty et al., 2014; Hemmat \& Baghi, 2018; Venuti et al., 2011; Zhang et al., 2018). Point-mutations in E5 may inhibit cell transformation, hinder activation of the cyclin A pathway, and disrupt the acidification of endosomes. The E5 protein of HPV has also been proven to increases the expression level of hepatocyte growth factor receptors, which promotes transformed cell invasiveness and regulates the proliferation of infected cells (Halavaty
et al., 2014; Hemmat \& Baghi, 2018; Venuti et al., 2011; Zhang et al., 2018). Conversely, studies have also demonstrated that the E5 ORF can tolerate many mutations and maintain the hydrophobic nature of E5, and that the conserved residues are sufficient to confer transforming activity (Mattoon et al., 2001; Venuti et al., 2011). As E5 is necessary for many pathways, minor sequence variations between E5 proteins from different HPV11 lineages and sub-lineages may significantly impact the function of this protein. Thus, functional analysis of E5 in HPV11 may further decipher the effects of E5 mutations on the HPV virus life cycle.

The URR is the most variable region within the HPV genome capable of accumulating and tolerating more nucleotide mutations as it does not encode proteins. The URR interacts with many cellular and viral factors and is involved with functions such as virus replication, gene expression, and transcription (Fang et al., 2020; Ribeiro et al., 2018). In the current study, the URR of 68 isolates was successfully sequenced, and genome analysis revealed 237 nucleotide substitutions, 118 nucleotide insertions and 90 nucleotide deletions. Isolates VBD34/08 and VBD28/14 had the most nucleotide differences compared to the A1 prototype M14119.1, with both containing 20 nucleotide substitutions and 12 nucleotide deletions. In addition, isolate VBD28/14 had the maximum number of nucleotide insertions of all isolates, 28, followed by 27 in VBD34/08.

A correlation between the transcriptional activity of HPV11 variants and RRP disease severity have previously been reported (Gáll et al., 2013). HPV11 presents E2 binding sites distributed along the URR regions and various promotors. These promotors are regulated differently, which indicates that independent regulation of early proteins is essential for the viral life cycle (DiLorenzo \& Steinberg, 1995; Dollard et al., 1992; Stoler et al., 1989). Duplication in the early viral promoter sequence of HPV11 has been associated with a higher degree of disease severity and the genome alterations T7904A and thymine at position 7546 , as in the A1 prototype sequence M14119.1, may have enhancer effects (Gáll et al., 2013). Due to the frequency of mutations in isolates from the current study, the functionality of binding sites in the URR may be impacted, affecting many cellular and viral factors. Although the URR is highly variable, the E2 transcription binding sites are reported to be highly conserved, and mutations are rarely found in these regions (Fang et al., 2020; Ribeiro et al., 2018).

Analysis of the E2 genome segment sequenced in this study revealed that most isolates (60/67) had only two point-mutations, and only a few (7/67) isolates had other mutations. The maximum number of point-mutations was five in VBD34/08 (2,4\%), followed by four in VBD28/14 (1,9\%). The E2 protein regulates essential factors during the viral life cycle, such as replication, transcription, and viral genome partitioning. Therefore, the binding of the E2 protein to corresponding binding sites is necessary for viral survival (Graham \& Faizo, 2017; Kardani \& Bolhassani, 2018a; Wallace \& Galloway, 2014).

Phylogenetic trees are a practical way to present the evolutionary relationship between organisms as the genetic sequence of contemporary sequences evolved from a common ancestor. Specific HPV11 genomic regions have previously been used to identify genetic variants as they provide sufficient genetic variation to differentiate between HPV11 variants. In 2011, Maver and colleagues identified two clades based on the topology of the maximum likelihood phylogenetic tree, namely the prototypic and non-prototypic clades (Maver et al., 2011). Subsequently, Burk and colleagues defined a new HPV sub-lineage as a $0.5 \%$ to $1.0 \%$ variation from the complete genome (Burk et al., 2011). This study revealed two clades and that the maximum pairwise difference between variants is approximately $0.4 \%$. They designated the two clades as sub-lineage A1 (previously termed prototypic variant group) and sub-lineage A2 (previously termed non-prototypic variant group) (Burk et al., 2011). Further comparative phylogenetic sequence analysis of isolates collected from Australia, Slovenia, China, Hungary, and Thailand revealed that the A2 sub-lineage predominates (Danielewski et al., 2013; Jelen et al., 2016). Results were further supported by a generally high degree of sequence conservation observed between diverse geographical regions throughout the world, suggesting that geographically specific variants were uncommon for HPV11 (Danielewski et al., 2013; Jelen et al., 2016; Maver et al., 2011). However, the discovery of new (sub)lineages in geographically restricted areas contradicts previous assumptions that intratypic variants of HPV11 are not geographically restricted (Jelen et al., 2016).

An extensive study on the genomic diversity of HPV11 from six continents revealed an additional lineage B and two additional sub-lineages A3 and A4, for the first time (Jelen et al., 2016). Lineage B had a $1.3 \%$ maximal pairwise distance, sub-lineage A4 had a complete nucleotide pairwise difference of above $0.5 \%$, and the two sub-lineage A 3 isolates had a total nucleotide pairwise difference in the range of $0.5 \%$ (Jelen et al., 2016). Results were visualised with a heatmap, and the specific complete nucleotide pairwise difference was not mentioned. A single, double peak was denoted as a wobble nucleotide in the LN833187 genome sequence submitted to GenBank, and therefore, the exact sequence is unknown (Appendix 4). As in the current study, phylogenetic analysis of E5a/b ORF, the L1 ORF and the URR using the maximum likelihood method identified sub-lineage A2 as the predominant sublineage globally (Jelen et al., 2016). For increased accuracy of phylogenetic analysis in the current study, 28 geographically distinct HPV11 representatives were included. Nonetheless, the tree topology generated in the present study did not correlate completely with that of the study published by Jelen and colleagues (Jelen et al., 2016).

Phylogenetic analysis of the concatenated L1-URR and E5a/b-L1-URR sequences identified in this study and sequences retrieved from GenBank correctly resolved into two lineages: Lineage A and lineage B. However, sub-lineage classification was frequently unclear. For example, isolates previously identified as sub-lineage A1 (FN907963.1) and A2 (FN907962.1), did not cluster closely with the
corresponding prototypes and instead, clustered together. Seven isolates from the current study clustered with isolates FN907963.1 and FN907962.1 with great certainty in the bootstrap test, therefore, phylogenetic analysis of the concatenated L1-URR and E5a/b-L1-URR is unable to reliably classify these isolates.

Neighbor-Joining and maximum likelihood trees were constructed using E2 segment data to determine whether E2 tree topology would be identical to phylogenetic trees constructed using concatenated E5a/b-L1-URR or L1-URR sequence data, and to determine whether the Neighbor-Joining tree and maximum likelihood tree based on E 2 segment sequence data would generate similar results. Maximum likelihood tree construction is computationally intensive, therefore, using the Neighbor-Joining method would be advantageous to generate results fast when using large data sets and for bootstrap analysis. In this study, the E2 segment tree topology was not identical to the phylogenetic trees using either the concatenated E5a/b-L1-URR or L1-URR. For instance, slight but essential variations were observed in the maximum likelihood tree constructed using the E2 segment. EU918768.1, FN870021.1 and FN907963.1 clustered correctly with M14119.1 when using the E2 segment, but not the concatenated E5a/b-L1-URR or L1-URR segments. However, in the study published by Jelen and colleagues (2016), analysis of the complete and partial genomes revealed EU918768.1, FN870021.1 and FN907963.1 clustered appropriately with the A1 prototype M14119.1.

In this study, we used phylogenetic software to calculate evolutionary distances between sequences and identify correspondence regions by computing the proportion of nucleotide differences between each pair of sequences. Analysis revealed isolates highly analogous to reference strains LN833169, LN833187.1, LN833165.1 and LN833183.1 appropriately clustered, and the results collated with those observed in the phylogenetic tree topology. Further analysis revealed that lineage B isolates VBD34/08 and LN833183.1 had the greatest divergence from the A1 prototype M14119.1, which is anticipated. Analysis of concatenated E5a/b-L1-URR revealed FN907962.1 as highly dissimilar, which is anomalous because FN907962.1 belongs to sub-lineage A2. LN833183.1 and VBD34/08 were the most different to LN833169.1 and LN833187.1, irrespective of the wobble nucleotide in the E2 segment. Analysis of estimates of evolutionary divergence between sequences revealed sub-lineage A2 isolates FR872717.1, JQ773408.1, JQ773411.1, JQ773412.1, and MK463914.1 were the most dissimilar to lineage B using concatenated E5a/b-L1-URR sequences and the E2 segment. However, the E2 segment also identified several other isolates belonging to various sub-lineages to be equally dissimilar.

Discrepancies observed in this study may be due to the large number of additional isolates included in this study or differences in software used to analyse the sequence data. Jelen and colleagues (2016) used RAxML and Bayesian trees to produce phylogenetic trees with complete genome data, and RAxML HPC2 (version 8.1.11) and MEGA (version 5) software to identify the most informative areas for whole-
genome-based phylogenetic clustering (Jelen et al., 2016). In the current study, MEGA X was used to analyse the most informative regions determined by Jelen and colleagues (2016).

Results obtained in the current study supports the notion of whole-genome sequencing for HPV11 classification below lineage level as a standard. Whole-genome sequencing generates more information regarding genetic variants and eliminates discrepancies seen in the current study. Concatenated L1URR, E5a/b-L1-URR sequences and the E2 segment had equivalent discriminatory power for lineage identification. However, concatenated E5a/b-L1-URR provides unclear results regarding sub-lineage classification. E2 Neighbor-Joining and maximum likelihood tree generated in this study generated a similar tree topology. Due to the inconsistencies and discrepancies observed in the phylogenetic tree topology using concatenated E5a/b-L1-URR, the sub-lineage classification of several isolates remained undetermined and therefore it is unclear whether tree topology using the E2 segment can reliably distinguish between sub-lineages. Investigations with a greater sample size are necessary to validate these findings.

It is essential to characterise the current strains circulating amongst the community to compare the various lineages and sub-lineages with disease severity in future studies, monitor the impact of the vaccination campaign on the circulating HPV types, and guide vaccine development. In addition, baseline data on circulating HPV genotypes may effectively monitor the impact of the vaccination program on the community. Also, determining the genotype of HPV responsible for RRP in patients may have prognostic implications.

# CHAPTER 3- Characterisation of novel human papillomavirus type 11 isolates 

### 3.1. Introduction

The identification and characterisation of potentially novel HPV lineages and sub-lineages has great public health significance (Sridhar et al., 2015). Little is known about the correlation between HPV lineages and sub-lineages and disease persistence, severity, prognosis, and malignant capability. The relevance of HPV variant classification has increased as HPV type and intratypic variants may possibly influence the course of disease (Gáll et al., 2013).

In the last two decades, NGS has been the most frequently used technique for identifying novel HPV types and variants. NGS refers to high-throughput sequencing and is capable of massive parallel sequencing of various DNA molecules (Ambulos et al., 2016; Kocjan et al., 2015). This relatively new technology is quickly becoming an indispensable tool in laboratories for both the detection and characterisation of clinically important pathogens such as HPV. NGS surpasses more conventional detection methods such as Sanger sequencing by faster turnaround time and the improved depth of knowledge regarding serotyping (Ambulos et al., 2016; Kocjan et al., 2015; Parker \& Chen, 2017). High-throughput sequencing is a powerful method that has been used to identify novel lineages and sub-lineages of HPV11 using a WGS approach. Complete genome sequencing using NGS and subsequent determination of percentage homology with known viral nucleotide sequences may aid in identifying novel lineages and sub-lineages (Sridhar et al., 2015).

The classification of HPV into genera, species and types based on the L1 ORF is well established. A L1 sequence with $>10 \%$ dissimilarity to any other deposited sequence is defined as a novel HPV type. A nucleotide variance of $0.5 \%-1.0 \%$ across the whole genome defines a variant sub-lineage, and a nucleotide variance of $1 \%-10 \%$ defines a variant lineage (Burk et al., 2011; Maver et al., 2011). However, the existence of a novel HPV variant is only confirmed when its complete genome is sequenced and deposited with the International HPV Reference Centre (https://www.hpvcenter.se/human_reference_clones), and the complete genome meets the classification requirements set by Burk and colleagues (2011).

Recently, one new lineage and two new sub-lineages of HPV11 were described, warranting further investigation into the genetics and distribution of these lineages (Jelen et al., 2016). Previous reports suggested that concatenated E5a/b-L1-URR sequences, as well as a 208bp gene segment in the E2 reportedly represent whole-genome variation and may be used to construct phylogenetic trees representative of the different lineages and sub-lineages (Jelen et al., 2016). Numerous isolates with
unique SNPs, amino acid changes and indels within the E5a/b, L1 and URR, as well as potential lineage $B$ isolates were identified using phylogenetic investigations. However sub-lineages could not be conclusively defined, suggesting that a novel lineage and sub-lineage requires a complete genome according to the classification and nomenclature system proposed previously (Bernard et al., 2010; Burk et al., 2013; Chen et al., 2011; de Villiers et al., 2004; Siqueira et al., 2016). Therefore, in the current chapter, whole genome sequences of selected isolates were determined using NGS and the genetic relationships analysed. In addition, the hypothesis that the partial genome sequences, as obtained in Chapter 2, generates tree topology which represents whole-genome variation is challenged (Jelen et al., 2016). Data on novel lineages and sub-lineages of HPV may contribute to future evolutionary- and vaccine studies, aid in vaccine development, and give insight into the pathogenicity of HPV11.

### 3.2. Materials and methods

### 3.2.1. Study samples

## Sample selection

Four isolates with unique SNPs, amino acid changes or indels within the E5a/b, L1 and URR were selected for whole-genome sequencing based on results from the phylogenetic analysis described in Chapter 2. Isolates which did not cluster closely with sub-lineage prototype sequences, as well as a possible lineage B isolate were selected for whole genome sequencing. Isolates included VBD28/14, VBD15/10, VBD74/09 and VBD01/10. Two isolates, VBD74/09 and VBD15/10, were isolated from one patient collected approximately five months apart to determine whether HPV11 experienced any nucleotide mutations between sample collection dates. Information on the selected isolates, including the date of birth, date of diagnosis, date of sample collection and sex, is summarised in Table 13. Informed consent for collection and storage of biopsies was obtained from each patient by Professor Seedat from the Department of Otorhinolaryngology, Faculty of the Health Sciences, UFS.

## HPV11 genome sequences retrieved from GenBank

Complete genome sequence data for 28 geographically distinct representatives of each HPV11, sublineages A1, A2, A3 and A4, and lineage B, were retrieved from GenBank (Appendix 2). This included the following isolates with GenBank accession numbers: M14119.1, LN833187.1, LN833185.1, LN833184.1, LN833183.1, LN833169.1, LN833165.1, LN833161.1, KU298879.1, JQ773412.1, JQ773411.1, JQ773409.1,JQ773408.1, JN644141.1, HE611263.1, HE574702.1, FR872717.1, FN907963.1, FN907962.1, FN870021.1, EU918768.1, MN788368.1, MK463921.1, MK463916.1, MK463914.1, MK313767.1, MK313765.1, and MK313763.1. Isolate M14119.1 was used as the sublineage A1 representative, LN833161.1 was used as the sub-lineage A2 representative, LN833169 was
used as the sub-lineage A3 representative, LN833187 was used as the sub-lineage A4 representative, and LN833183 was used as the lineage B representative. Sequence length ranged from 7932bp to 7949bp and is summarised in Appendix 2.

Ethical clearance was obtained from the Health Sciences Research Ethics Committee of the UFS (UFSHSD2019/1109/2708) (Appendix 3).

### 3.2.2. Full-genome amplification and sequencing

The complete genome was determined by amplifying two overlapping DNA fragments representing the complete HPV11 genome. Primers used for amplification are summarised in Table 11 (Jelen et al., 2016). Visual representation of overlapping fragments is depicted in Figure 8. For both overlapping PCR fragments, a Phusion ${ }^{\circledR}$ High-Fidelity PCR kit (Thermo Scientific, Illinois, USA) was utilised according to the manufacturer's instructions. Contents of the reaction mixture are summarised in Table 12. The cycling conditions were carried out as follows: $98^{\circ} \mathrm{C}$ for two minutes, followed by 40 cycles of $98^{\circ} \mathrm{C}$ for 10 seconds, $53.5^{\circ} \mathrm{C}$ for 30 seconds, and $72^{\circ} \mathrm{C}$ for three minutes. The final extension step was carried out at $72^{\circ} \mathrm{C}$ for eight minutes before cooling to $4^{\circ} \mathrm{C}$.

Table 11: List of primers used for sequencing full-length human papillomavirus type 11 (HPV11) genome in two overlapping fragments.

| Region | Primer name | Primer sequence $5^{\prime} \rightarrow 3^{\prime}$ | $\begin{gathered} \text { GC } \\ \text { content }^{\text {c }} \end{gathered}$ | Tm ${ }^{\text {c }}$ | Primer binding site ${ }^{\text {b }}$ | $\begin{aligned} & \text { Amplicon } \\ & \text { size }^{d} \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $3^{\prime}$ end of the E2 ORF, the E5a/b ORF, the L2 ORF, the L1 ORF, the URR, and the $5^{\prime}$ end of the E6 ORF | $\text { HPV-11-S2F }{ }^{\text {a }}$ $\text { HPV-11-S2R }{ }^{\text {a }}$ | TTACAACAAGC ACCAAAGAAG <br> TTCTATTTCACA CAACGGCT | $\begin{aligned} & 38.10 \% \\ & 40.00 \% \end{aligned}$ | $\begin{aligned} & 60.1^{\circ} \mathrm{C} \\ & 60.7^{\circ} \mathrm{C} \end{aligned}$ | $\text { nt } 3529 \text { to } 3549$ <br> nt 427 to 446 | 4851bp |
| $3^{\prime}$ end of the L1 ORF, the URR, the E6 ORF, the E7 ORF, the El ORF, and a segment of the E2 ORF | HPV-11- $\text { S1MS- } \mathrm{F}^{\mathrm{a}}$ <br> HPV-11- <br> S1MS-R ${ }^{\text {a }}$ | GGATATGAGTTT <br> TTGGGAGGT <br> ATGCCACGTTGA AGATGCTA | $\begin{aligned} & 42.86 \% \\ & 45.00 \% \end{aligned}$ | $\begin{aligned} & 60.7^{\circ} \mathrm{C} \\ & 62.9^{\circ} \mathrm{C} \end{aligned}$ | nt 7084 to 7104 <br> nt 3660 to 3679 | 4,521bp |
| ${ }^{\text {a }}$ Primers reported by Jelen et al., 2016. <br> ${ }^{b}$ Positions of the nucleotides determined with respect to the prototype human papillomavirus type 11 genome (GenBank acc. no. M14119). <br> ${ }^{\text {c }}$ Melting temperature (Tm) and GC content determined by Thermo Fisher Scientific Tm calculator (Allawi \& SantaLucia, 1997). <br> ${ }^{d}$ Size determined with respect to the prototype human papillomavirus type 11 genome (GenBank acc. no. M 14119). |  |  |  |  |  |  |

Table 12: Polymerase chain reaction (PCR) components for amplifying human papillomavirus type 11 full-length genome in two overlapping fragments (E1 to L1 genes; L1 to E1 genes).

| Component | Volume | Final concentration |
| :--- | :--- | :--- |
| Phusion DNA polymerase | $0,5 \mu \mathrm{l}$ | 1,0 units |
| 5 x Phusion HF Buffer | $10 \mu \mathrm{l}$ | 1 X |
| $10 \mu \mathrm{M}$ Forward primer | $2,5 \mu \mathrm{l}$ | $0,5 \mu \mathrm{M}$ |
| $10 \mu \mathrm{M}$ Reverse primer | $2,5 \mu \mathrm{l}$ | $0,5 \mu \mathrm{M}$ |
| 10 mM dNTPs | $1 \mu \mathrm{l}$ | $200 \mu \mathrm{M}$ |
| Template DNA | $5 \mu \mathrm{l}$ | $\mathrm{N} / \mathrm{A}$ |
| Nuclease free water | $27,5 \mu \mathrm{l}$ | $\mathrm{N} / \mathrm{A}$ |
| Total | $50 \mu \mathrm{l}$ | $\mathrm{N} / \mathrm{A}$ |



Figure 8: Genome organisation of a low-risk human papillomavirus type 11 (HPV11).
E1-E7 early genes, L1-L2 late genes, URR Upper regulatory region, AE early polyA signal, and P97 and P742 promotors are indicated. The figure is drawn based on HPV11 A1 prototype (GenBank accession number M14119.1). The yellow arrow line indicates genome fragment amplified by primer pair HPV-11-S2F and HPV-11-S2R, and the black arrow line indicates genome fragment amplified by the HPV-11-S1MS-F and HPV-11-S1MS-R primer pair.

### 3.2.3. Agarose gel electrophoresis

PCR amplicons were separated and visualised by electrophoresis using a $1 \%$ agarose gel. Briefly, a $1 \%$ gel was prepared using 1 g Seakem® LE agarose powder, and 100 ml of 1 x TAE buffer at pH 8.0 and the gel was electrophoresed at 90 V for 45 minutes. Tables summarising the preparation of TAE and $1 \%$ agarose gels is included in Appendix 6 and 7. The DNA was stained using a GelRed nucleic acid gel stain (Thermo Fisher Scientific, USA) (Appendix 8) and visualised using the BioRad Molecular Imager Gel Doc ${ }^{\text {TM }}$ XR+ with Image Lab ${ }^{\text {TM }}$ Software (BioRad, California, USA) to determine the fragment sizes according to the known DNA size marker. The O'GeneRuler ${ }^{\text {TM }}$ 100bp DNA ladder SM0333 (Fermentas, Illinois, USA) containing DNA fragments from 100bp to 10 000bp was used to estimate the size of the amplicons.

### 3.2.4. Purification of PCR product

The Wizard® SV Gel and PCR Clean-up System (Promega, Madison, USA) was used to purify target DNA from a $1 \%$ agarose gel according to manufacturer's instructions. Membrane Binding Solution was added at a ratio of $10 \mu \mathrm{l}$ of solution per 10 mg of agarose gel and the gel heated at $60^{\circ} \mathrm{C}$ until completely dissolved. The contents were transferred to the SV Minicolumn and washed twice by adding $700 \mu \mathrm{l}$ and $500 \mu 1$ of Membrane Wash Solution. The DNA was eluted in $50 \mu 1$ of nuclease-free water and stored at $4^{\circ} \mathrm{C}$ until further use.

### 3.2.5. Determination of DNA concentration

DNA concentration was measured using the Qubit fluorometer and the Qubit dsDNA BR assay kit, according to the manufacturer's instructions (Thermo Fisher Scientific, Illinois, USA). A working solution was made by diluting dsDNA BR reagent and BR buffer (1:200) in microcentrifuge tubes. Then, $190 \mu \mathrm{l}$ working solution and $10 \mu \mathrm{l}$ of each standard were loaded in microcentrifuge tubes. In separate microcentrifuge tubes, $198 \mu$ l working solution and $2 \mu 1$ sample were loaded, vortexed and incubated for two minutes at room temperature. The Qubit was calibrated, and DNA concentrations of all samples were measured.

### 3.2.6. MiSeq library preparation and sequencing

The Nextera XT DNA Library Preparation kit (Illumina, California, USA) was used to convert the purified DNA product to a short, fragmented DNA library, followed by size selection using AMpure XP beads (Beckman Coulter, California, USA). The multiplexed libraries were analysed on an Illumina MiSeq (Illumina, California, USA) with the MiSeq reagent kit v3 (300 cycles) (Illumina, California, USA) at the UFS Next Generation Sequencing Unit.

### 3.2.7. Next-generation sequencing data analysis

FASTQ sequences were imported into Geneious 2021 software (https://www.geneious.com) as paired ends with an insert size of 500 . Reads were trimmed with an error probability limit of 0.05 and filtered. Reads were assembled and aligned to the A1 prototype M14119.1 using a medium sensitivity, and a consensus sequence was generated using the highest quality threshold. Using Geneious 2021 software (https://www.geneious.com), the consensus sequence was inspected to identify ambiguities and degenerative bases and exported as text documents. An assembly report was generated and the length of the genome, CG content and pairwise identity was determined using Geneious 2021 software.

## Analysis of nucleotide and amino acid variation

Genomic variants and genomic positions were identified by comparing the sequence data with the prototype HPV11 genome (GenBank accession number M14119.1). A corrected sequence with a 2bp insertion at genomic position 7717-7718 was used when determining genomic variants and genomic positions (Maver et al., 2011). For the isolates sequenced in this study, nucleotide substitutions, nucleotide insertions, and nucleotide deletions were identified with respect to the A1 prototype strain (GenBank accession number M14119.1) using MEGA X software (Kumar et al., 2018).

The predicted amino acid sequences for each coding region were determined with the use of a codon chart included in Appendix 9. Non-synonymous substitutions resulting in amino acid alterations were identified by comparing amino acid data with the HPV11 A1 prototype (GenBank accession number M14119.1).

## Percentage variation in coding regions

To estimate the percentage of variation within the coding regions, the number of nucleotide- and amino acid variations compared to the A1 prototype sequence M14119.1 were converted into a percentage using the following equations:
$\frac{\text { Number of nucleotide variances within a spesific coding region compared to M14119.1 }}{\text { Number of nucleotides within spesific coding region of M14119.1 }} \times 100$
and
$\frac{\text { Number of amino acid variances within a spesific coding region compared to M1411.1 }}{\text { Number of amino acids within spesific coding region of M14119.1 }} \times 100$

## Degree of divergence

The degree of divergence was calculated using pairwise analysis of percentage divergence of nucleotides with MEGA X software using the Maximum Composite Likelihood model. Codon positions
included were $1^{\text {st }}, 2^{\text {nd }}, 3^{\text {rd }}$, and non-coding. All positions with less than $95 \%$ site coverage were removed using the partial deletion option. The degree of divergence was calculated for isolates sequenced in the current chapter, as well as the 28 complete genomes retrieved from GenBank and divergence from lineage A and B prototypes were described. The degree of divergence between isolates sequenced in the current chapter was also calculated.

## Maximum likelihood method

The evolutionary history was inferred using the maximum likelihood method and the Tamura-Nei model with a bootstrap value of 1000 software using sequence data determined in the current chapter, and 28 complete sequences retrieved from GenBank. Analysis was performed using MEGA X software. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, then selecting the topology with the higher log-likelihood value. The $1^{\text {st }}, 2^{\text {nd }}, 3^{\text {rd }}$, and non-coding codon positions were included in the analysis.

Isolates with known (sub)lineage classification based on phylogenetic analysis were aligned with the corresponding reference strain using MEGA X software to identify nucleotide- and amino-acid variabilities (Kumar et al., 2018).

### 3.3. Results

### 3.3.1. $\quad$ Selection of isolates for whole genome sequencing

Four isolates were selected for complete genome determination. Two isolates (VBD15/10 and VBD74/09) from one patient collected approximately five months apart were selected because the concatenated E5a/b-L1-URR nucleotide sequence profiles were highly similar and therefore considered representative of two isolates retrieved from GenBank for which compete genome data was available and eight isolates from the current study of RRP patients. VBD01/10 was selected because the concatenated E5a/b-L1-URR nucleotide sequence profiles were highly similar to an isolate retrieved from GenBank for which compete genome data was available and representative of six isolates from the current study. These isolates did not cluster closely with any prototype sequence. VBD34/08 and VBD28/14 were identified as potential lineage B isolates based on the tree topology however, there was insufficient DNA available to obtain further sequence data for VBD34/08. Thus, overall, the complete genome of four HPV11 isolates, VBD01/10, VBD15/10, VBD28/14 and VBD74/09, from three patients were sequenced in this study.

Information regarding patient sex, date of birth, date of diagnosis and sample collection is included in Table 13. The DNA sequencing reads were mapped to the A1 prototype M14119.1 and a consensus
sequence was generated for each isolate. The number of assembled reads is summarised in Table 14. The minimum and maximum lengths of the consensus sequences, the CG content and the pairwise identity are summarised in Table 15.

Additional genome sequence data for 28 geographically distinct representatives of each HPV11 lineage, including lineage A, sub-lineages A1 to A4, and lineage B, were retrieved from GenBank and used for downstream analysis (Appendix 2).

Table 13: Patient information.

| Isolate | Sex | Date of birth | Date of diagnosis | Date of sample collection |
| :---: | :---: | :---: | :---: | :---: |
| VBD01/10 | Male | $2007 / 01 / 04$ | $2010 / 01 / 06$ | $2009 / 11 / 03$ |
| VBD28/14 | Male | $2009 / 03 / 29$ | $2014 / 06 / 26$ | $2010 / 01 / 06$ |
| VBD74/09 | Male | 2009/10/13 | $2009 / 10 / 23$ | $2009 / 11 / 03$ |
| VBD15/10 | Male |  |  | $2014 / 06 / 26$ |

Table 14: Next-generation sequencing assembly report for four complete human papillomavirus type 11 (HPV11) mapped to the A1 prototype M14119.1 using Geneious Prime software (https://www.geneious.com).

|  | Assembly report |  |  |
| :---: | :---: | :---: | :---: |
| Isolate | Total reads | Assembled reads | Not assembled reads |
| VBD01/10 | $1,033,592$ | $1,022,153$ | 11,439 reads |
| VBD15/10 | $1,652,362$ | $1,636,151$ | 16,211 reads |
| VBD28/14 | $1,333,518$ | $1,306,323$ | 27,195 reads |
| VBD74/09 | $1,903,068$ | $1,879,877$ | 23,191 reads |

Table 15: The lengths of the complete human papillomavirus type 11 (HPV11) consensus sequence, the CG content and the pairwise identity of four HPV11 isolates mapped to the Al prototype M14119.1 using Geneious Prime software (https://www.geneious.com).
The corrected M14119.1 with a two base-pair (bp) insertion at genomic position 7717-7718 was used (Maver et al., 2011).

| Isolate | Sequence length | CGcontent | Pairwise identity |
| :--- | :--- | :--- | :--- |
| VBD01/10 | 7934 bp | $41.1 \%$ | $99.3 \%$ |
| VBD15/10 | 7932 bp | $41.1 \%$ | $99.4 \%$ |
| VBD28/14 | 7949 bp | $41.1 \%$ | $99.3 \%$ |
| VBD74/09 | 7932 bp | $41.1 \%$ | $99.4 \%$ |

### 3.3.2. Nucleotide and amino acid variances across the human papillomavirus type 11 genome

The nucleotide sequence variabilities for the four complete HPV11 genomes are summarised in Table 16. Briefly, VBD01/10 contained 25 nucleotide substitutions, a nucleotide deletion and two nucleotide insertions compared to the A1 prototype M14119.1, VBD15/10 and VBD74/09 contained 26 nucleotide substitutions and one nucleotide deletion compared to the A1 prototype M14119.1, and VBD28/14 contained 94 nucleotide substitutions, 12 nucleotide deletions and 28 nucleotide insertions compared to the A1 prototype M14119.1. In total, 45 indels were identified among four isolates in the URR.

Nucleotide insertions include an ACGCGC insertion following genomic position 7529 in VBD28/14, a C and CC insertion following genomic position 7584 in VBD28/14 and VBD01/10, respectively, a CGCCAGG insertion following genomic position 7723 in VBD28/14, and a GGTTGTGTTATCTC insertion following genomic position 7746 in VBD28/14. Nucleotide deletions include a T deletion at genomic position 7509 in VBD01/10, VBD15/10 and VBD74/09, and a GTATCTTGCCAA deletion from genomic position 7694 to 7705 in VBD28/14. All isolates contained nucleotide substitutions T137C, C1783G, G1784C, A2580G, C2884T, T2888C, G3436A, A3645G, A3832G, A3952T, G3991C, C4647T, C4887A, and T7547C. VBD15/10 and VBD74/09 had identical nucleotide substitutions, insertions, and deletions.

The amino acid sequence variabilities for the four complete HPV11 genomes are summarised in Table 17. By aligning the four isolates sequenced in the current study and the A1 prototype sequence (GenBank accession number M14119.1), 69 non-synonymous substitutions resulting in changed amino acids were identified. Briefly, VBD01/10 had ten non-synonymous substitutions, VBD15/10 and VBD74/09 had 12 non-synonymous substitutions, and VBD28/14 had 35 non-synonymous substitutions. All isolates contained an arginine ( R ) to A substitution in the E1 ORF, a K to R nonsynonymous substitution in the E2 ORF, a G to E non-synonymous substitution in the E4 ORF, and a valine to F and T to L in the E5a ORF. VBD15/10 and VBD74/09 had identical amino acid sequences.

As previously mentioned, a nucleotide variance of $0.5 \%-1.0 \%$ across the whole genome defines a variant sub-lineage and the nucleotide sequence of the L1 ORF is often used to represent the whole genome variation due to the high degree of conservation in this region (Burk et al., 2011, 2013). Subsequently, to estimate the degree of variation in coding regions, the percentage difference between isolates in the current study and the A1 prototype sequence M14119.1 was determined. In brief, the nucleotide sequences of coding regions varied from $0.46 \%$ in the L1 ORF to $2.22 \%$ in the E5b ORF. The amino acid sequences variability ranged from $0.39 \%$ in the L1 ORF to $6.75 \%$ in the E5b ORF.

Table 16: Sequence comparison of nucleotide variances among four human papillomavirus type 11 (HPV11) isolates against the HPV11 prototype with GenBank accession number M14119.1.
Residues that differed from those of the HPV11 prototype are indicated by a white (nucleotide substitutions), green (nucleotide deletions) and yellow (nucleotide insertions) colour change and the substituted-, deleted- or inserted nucleotides. The corrected M14119.1 with the 2bp insertion at genomic position 7717-7718 was used (Maver et al., 2011).


Table 17: Comparison of amino acid variances among four human papillomavirus type 11 (HPV11) isolates against the HPV11 prototype with GenBank accession number M14119.1.
The total number of amino acids in each gene and the amino acid number where a non-synonymous is observed is indicated. Residues that differed from those of the HPV11 prototype are indicated by a white colour difference and the substituted amino acid.

| Gene | Amino acid no. | Amino acid* | Isolate |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | VBD01/10 | VBD15/10 | VBD28/14 | VBD74/09 |
| E6 | 65 | A |  |  | V |  |
| $1 . .150$ | 122 | G |  |  | E |  |
| E7 | 12 | V |  |  | L |  |
| $1 . .98$ | 45 | A | S | S |  | S |
| $\begin{gathered} \mathrm{E} 1 \\ 1 . .649 \end{gathered}$ | 100 | N |  | T |  | T |
|  | 174 | A |  |  | A |  |
|  | 184 | D |  |  | N |  |
|  | 318 | R | A | A | A | A |
|  | 527 | Q |  |  | H |  |
| $\begin{gathered} \mathrm{E} 2 \\ 1 . .367 \end{gathered}$ | 247 | N |  | T |  | T |
|  | 257 | S | F |  |  |  |
|  | 263 | N |  |  | K |  |
|  | 265 | I |  |  | S |  |
|  | 272 | K |  |  | T |  |
|  | 282 | A |  |  | T |  |
|  | 283 | A |  |  | D |  |
|  | 308 | K | R | R | R | R |
|  | 350 | K |  | N |  | N |
| $\begin{array}{\|c} \mathrm{E} 4 \\ 1 . .108 \end{array}$ | 46 | Q | R |  |  |  |
|  | 61 | G | E | E | E | E |
|  | 70 | T |  | P |  | P |
|  | 78 | T | L | L |  | L |
|  | 80 | D | S |  |  |  |
|  | 86 | T |  |  | K |  |
|  | 88 | S |  |  | A |  |
|  | 95 | S |  |  | R |  |
|  | 106 | L |  |  | I |  |
| $\begin{aligned} & \text { E5A } \\ & 1 . .91 \end{aligned}$ | 11 | I |  |  | G |  |
|  | 28 | V | F | F | F | F |
|  | 41 | T | L | L | L | L |
|  | 91 | Q |  | P |  | P |
| $\begin{aligned} & \text { E5B } \\ & 1 . .74 \end{aligned}$ | 7 | H |  |  | Q |  |
|  | 28 | L |  |  | F |  |
|  | 43 | K |  |  | N |  |
|  | 56 | T |  |  | N |  |
|  | 63 | S |  |  | L |  |
| $\begin{gathered} \mathrm{L} 2 \\ 1 . .455 \end{gathered}$ | 191 | D |  |  | E |  |
|  | 215 | A |  |  | T |  |
|  | 292 | L |  |  | I |  |
|  | 325 | V |  |  | I |  |
|  | 334 | L |  |  | M |  |
|  | 351 | F |  |  | L |  |
|  | 368 | L |  |  | F |  |
| L1 | 176 | S |  |  | T |  |
| 1.. 501 | 476 | A |  |  | V |  |
| *A - alanine, D - aspartic acid, E - glutamic acid, F - phenylalanine, G - glycine, H - histidine, I - is oleucine, K - lysine, L - leucine, M - methionine, N - asparagine, P - proline, Q glutamine, R - arginine, S - serine, T - threonine, V - valine |  |  |  |  |  |  |

### 3.3.3. Pairwise nucleotide difference between complete human papillomavirus type

 11 genomesPairwise nucleotide difference between complete genomes obtained in this study and sequence data retrieved from GenBank is summarised in Appendix 12. There were a total of 7920 positions in the final dataset. Pairwise nucleotide difference between the four complete genomes obtained in this study and 28 complete genome sequences obtained from GenBank varied from 0.000 to 0,01275 (Appendix 12). Divergence from the sub-lineage A1 prototype M14119.1 ranged from 0,00038 (FN907963.1) to 0,01224 (LN833183.1) with an average divergence of 0,00378 . Divergence from the sub-lineage A2 prototype LN833161.1 ranged 0,00025 (HE611263.1 and JQ773409.1) to 0,01224 (LN833183.1) with an average divergence of 0,00241 . Divergence from the sub-lineage A3 prototype LN833169.1 ranged from 0,00139 (LN833165.1) to 0,01275 (LN833183.1) with an average divergence of 0,00578 . Divergence from the sub-lineage A4 prototype LN833187.1 ranged from 0,00507 (LN833165.1) to 0,01275 (LN833183.1) with an average divergence of 0,00624 . Divergence from the lineage B prototype LN833183.1 ranged from 0,00051 (VBD28/14) to 0,01275 (LN833169.1, LN833187.1) with an average divergence of 0,01186 (Appendix 12).

Pairwise nucleotide difference between the four complete genomes obtained in this study is summarised in Table 18. There were a total of 7920 positions in the final dataset. VBD15/10 and VBD74/09 had an evolutionary divergence of zero.

Table 18: Estimates of evolutionary divergence between four human papillomavirus type 11 (HPV11) sequences using MEGA X software.
The corrected M14119.1 with the 2bp insertion at genomic position 7717-7718 was used (Maver et al., 2011).

|  | VBD01/10 | VBD15/10 | VBD28/14 | VBD74/09 |
| :--- | ---: | ---: | ---: | ---: |
| VBD01/10 |  |  |  |  |
| VBD15/10 | 0,00139 |  |  |  |
| VBD28/14 | 0,01184 | 0,01197 |  |  |
| VBD74/09 | 0,00139 | 0 | 0,011973 |  |

### 3.3.4. Maximum Likelihood method

The evolutionary history was inferred by using the maximum likelihood method and Tamura-Nei model. The tree with the highest log likelihood (-12751.59) is depicted. The analysis involved 32 nucleotide sequences and 7972 positions were included in the final dataset. Phylogenetic analysis of the HPV11 complete genome resolved into two lineages, namely lineage A and lineage B.

Two isolates, VBD28/14 and LN833183, (2/32) were identified as lineage B and clustered together with a $100 \%$ certainty in the bootstrap test (Figure 9). The majority of isolates $(23 / 28,82 \%)$, including VBD01/10, VBD15/10, and VBD74/09, clustered with the A2 reference LN833161.1. VBD15/10 and VBD74/09 clustered separately from the other sub-lineage A2 isolates (Figure 9).
3.3.5. Genome heterogeneities with respect to the corresponding prototype

Inferred by phylogenetic analysis, VBD28/14 was identified as a lineage B isolate. There were three nucleotide differences identified between VBD28/14 and the lineage B prototype LN833183.1. Nucleotide heterogeneities are summarised in Table 19. Based on the alignment of VBD28/14 with the lineage B prototype LN833183.1, one non-synonymous substitution could be identified. A C6831T nucleotide substitution resulted in an A to E amino acid substitution (Table 19).

Table 19: Genomic variations between human papillomavirus type 11 (HPV11) lineage B isolates VBD28/14 sequenced in this study and lineage B prototype LN833183.1 retrieved from GenBank.
The corrected M14119.1 with the 2bp insertion at genomic position 7717-7718 was used (Maver et al., 2011).

|  | Genomic position |  |  |
| :---: | :---: | :---: | :---: |
| Isolate | 48 | 3820 | 6831 |
| VBD28/14 | T | A | C |
| LN833183.1 | C | G | T |

VBD01/10, VBD15/10 and VBD74/09 were identified as sub-lineage A2 isolates through phylogenetic analysis. Based on the alignment of VBD01/10, VBD15/10 and VBD74/09 with the A2 prototype LN833161.1, a total of 14 variable nucleotide positions and nine variable amino acid positions were identified. Nucleotide heterogeneities are summarised in Table 20.

Table 20: Genomic variations between human papillomavirus type 11 (HPV11) sub-lineage A2 isolates sequenced in this study and sub-lineage A2 prototype LN833161.1 retrieved from GenBank.
The corrected M14119.1 with the $2 b p$ insertion at genomic position 7717-7718 was used (Maver et al., 2011).

|  | Genomic position |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Isolate | 1107 | 1130 | 1623 | 2358 | 3391 | 3462 | 3492 | 3772 | 4142 | 4380 | 7252 | 7272 | 7413 | 7585-7586 |
| LN833161.1 | C | A | T | T | G | A | C | A | A | T | G | C | C | C insertion |
| VBD01/10 |  |  |  |  |  |  | T |  |  |  | A | A |  | CC insertion |
| VBD15/10 | A | C | C | C | A | C |  | C | C | G | A | A | A | No insertion |
| VBD74/09 | A | C | C | C | A | C |  | C | C | G | A | A | A | No insertion |

One non-synonymous substitution was identified in the E1 ORF of two isolates. An A1130C nucleotide substitution resulted in a N to T in VBD15/10 and VBD74/09 at amino acid 100. Two non-synonymous substitutions were identified in the E2 ORF in both VBD15/10 and VBD74/09. An A3462C nucleotide substitution resulted in a N to T amino acid substitution at amino acid 247, and an A3772C nucleotide substitution resulted in a K to N amino acid substitution at amino acid 350 . One non-synonymous substitution was identified in the E2 ORF in VBD01/10. A C3492T nucleotide substitution resulted in a S to F amino acid substitution at amino acid 257 . VBD15/10 and VBD74/09 both contained two nonsynonymous substitutions in the E4 ORF. A G3391A nucleotide substitution resulted in a R to Q amino acid substitution at amino acid 46 and an A3462C nucleotide substitution resulted in a T to P amino acid substitution at amino acid 70. One non-synonymous substitution was identified in the E4 ORF in VBD01/10. A C3492T nucleotide substitution resulted in a P to S amino acid substitution at amino acid 80. One non-synonymous substitution was observed in VBD15/10 and VBD74/09 in the E5a ORF. An A4142C nucleotide substitution resulted in a Q to P amino acid substitution at amino acid 91 . One nonsynonymous substitution was identified in three isolates in the L1 ORF. A C7272A nucleotide substitution resulted in a T to K amino acid substitution at amino acid 501 in VBD01/10, VBD15/10 and VBD74/09 (Table 20).


Figure 9: Phylogenetic tree of human papillomavirus type 11 (HPV11) based 32 nucleotide sequence alignments of HPV11 including four isolates sequenced in this study and 28 sequences of geographically distinct isolates retrieved from GenBank. M14119.1 (blue) was used as a representative of sub-lineage A1, LN833161.1 (green) as a representative of sub-lineage A2, LN833169.1 (yellow) as a representative of sub-lineage A3, LN833187.1 (red) as a representative of sub-lineage A4 and LN833183.1 (orange) was used as a representative of lineage B. The maximum likelihood tree was constructed using MEGA $X$ using the maximum likelihood method and Tamura-Nei model. The corrected M14119.1 with the $2 b p$ insertion at genomic position 7717-7718 was used (Maver et al., 2011).

### 3.4. Discussion

HPV classification below types is increasingly relevant since within-type variants (lineages and sublineages) are associated with differential prognosis and outcomes concerning virus persistence, development of lesions and progression to malignancy (Flores-Díaz et al., 2017). The advent of metagenomics using high-throughput Illumina sequencing has enabled the discovery of numerous novel viruses in various bio-niches at a reasonable cost and unprecedented speed. Additionally, genome characterisation assists in studying viral diversity and the association of virus and disease (Ambulos et al., 2016; Kocjan et al., 2015; Parker \& Chen, 2017; Tuna \& Amos, 2017).

The HPV nomenclature is well established. However, while classification into species and types is based on unique genes, variant lineage and sub-lineage classification are based on the complete genome of the virus. HPV variants are considered novel only after the complete genome has been cloned and deposited with the International HPV Reference Centre (Bernard et al., 2010; Burk et al., 2011; Bzhalava et al., 2015; de Villiers et al., 2004; Kocjan et al., 2015) (https://www.hpvcenter.se/human_reference_clones). Phylogenetic analysis of concatenated E5a/b-L1URR sequences suggested a supposed novel HPV11 lineage or sub-lineage and isolates of interest using a previously published PCR protocol (Jelen et al., 2016; Maver et al., 2011). However, due to the discrepancies observed in the phylogenetic tree topology using concatenated E5a/b-L1-URR, the sublineage classification of several isolates remained undetermined in Chapter 2. It was also unclear whether the E2 segment can reliably distinguish between sub-lineages. Therefore, the complete HPV11 viral genome of selected isolates was amplified using long-range PCR and sequenced using highthroughput sequencing.

Isolates characterised in this study were initially isolated from respiratory tract papilloma samples obtained from patients with a clinical diagnosis of RRP. We assembled four HPV11 complete genomes from three different patients. VBD28/14 had the most nucleotide substitutions and indels, followed by VBD01/10. VBD15/10 and VBD74/09, isolated from one patient, were identical and indicated no signs of evolution within the host. Comparison of the ORFs of each of the four isolates identified nonsynonymous substitutions that changed the amino acid, which may alter protein function and expression, which underlie disease phenotype (Lebeuf-Taylor et al., 2019). Also, interference in gene expression may lead to cancer progression and failure of treatment options. Therefore, future studies regarding viral gene expression are vital.

Phylogenetic analysis confirmed that VBD28/14 clusters with another lineage B isolate which was recently identified (Jelen et al., 2016), and analysis of the genome organisation of VBD28/14 revealed many similarities and some differences, including three nucleotide substitutions and one nonsynonymous substitution in the L1 capsid protein. VBD15/10 and VBD74/09 was identified as sub-
lineage A2 but clustered separately from all other sub-lineage A2 isolates. VBD15/10 and VBD74/09 had seven non-synonymous heterogeneities in the E1-, E2-, E4-, E5a- and L1 ORF compared to the A2 prototype used. VBD01/10 was the most similar to the A2 prototype and only had three nucleotide substitutions, one nucleotide insertion and three non-synonymous substitutions in the E2, E4 and L1 ORFs.

Conformational changes in the E1 and E2 proteins may affect replication of the HPV11 circular dsDNA genome. Furthermore, E2 possesses an initiation codon for E4, and protein changes in either may affect virion release and apoptosis in terminally differentiated cells (Graham \& Faizo, 2017; Kardani \& Bolhassani, 2018a; Wallace \& Galloway, 2014). Point-mutations in E5 may inhibit cell transformation, hinder activation of various pathways, disrupt the acidification of endosomes, promote transformed cell invasiveness, and misregulate the proliferation of infected cells. As E5 is necessary for many pathways, sequence variations may significantly impact the function of this protein (Halavaty et al., 2014; Hemmat \& Baghi, 2018; Venuti et al., 2011; Zhang et al., 2018). Functional analysis of early proteins in HPV11 variants may further decipher the effects of mutations on the HPV virus life cycle.

Within the L1 ORF, sequence variation was low, likely due to the slow genomic evolutionary rate (Chen et al., 2009, 2011). However, genomic- and amino acid variances within the L1 were identified in all samples. In order to determine whether non-synonymous nucleotide substitutions have an effect on the biological function of vaccines, functional analysis is required. Amino acid- and protein alterations that occur within the conserved epitopes of L1, have the ability to influence vaccine efficacy. However, to date, there have been no reports suggesting that the biological function of HPV type specific vaccines is influenced by antigenic differences between intratypic variants (Barra et al., 2019; Benedict \& Derkay, 2021; Buck et al., 2013; Chabeda et al., 2018; Dadar et al., 2018). The consequences of polymorphisms in the novel lineage B, isolates of interest VBD15/10 and VBD74/09, and sub-lineage A3 and A4 on antibody recognition following natural infection or vaccination by the present-day vaccines and effects on the viral life cycle requires additional investigation.

In conclusion, during this study the complete genome of four isolates of HPV11 were successfully identified and characterised from three patients. Additionally, this is also the second report of a lineage $B$ isolate identified. Both lineage B isolates identified thus far have been isolated in South Africa, suggesting that this lineage B may be geographically restricted or have a low infectivity rate. Further research on sub-lineage A2 isolates, VBD15/10 and VBD74/09, is necessary to characterise these isolates fully. Research with different sample types and larger sample sizes are needed to clarify the prevalence and disease association and the impact of those newly characterised variants on HPV persistence and RRP development.

## Chapter 4 - Concluding remarks

Papillomaviruses have a well-documented history with human hosts, and most humans are infected at some point in their lives. Papillomaviruses have a slow evolutionary rate of approximately $10^{-8}$ nucleotide substitutions per site per year and experience little recombination during their evolution; hence nucleotide polymorphisms mainly transpire due to arbitrary mutation time (Bernard, 1994; Chen et al., 2009, 2011; van Doorslaer, 2013). HPV was discovered in 1949 (Strauss et al., 1949), and evidence suggesting the involvement of HPV11 in RRP was first reported in 1982 (Gissmann et al., 1982). To date, over 220 HPV types are recognised globally (https://www.hpvcenter.se/human_reference_clones).

Despite phylogenetic relatedness, knowledge regarding intratypic variants of HPV11 and association with the clinical outcomes of infection are limited. Genetic traits resulting in various lineages and sublineages of the virus are all encoded within the relatively small eight kilobyte circular dsDNA genome. HPV11 is ordinarily associated with a more severe disease outcome in patients with RRP and earlier age at diagnosis compared to HPV6, suggesting increased virulence and pathogenesis (El Achkar et al., 2020; Intakorn \& Sonsuwan, 2014; Omland et al., 2014; Seedat, 2020).

According to previous South African surveillance studies, between $64 \%$ and $71 \%$ of the population are infected with HPV, and between $1 \%$ and $4 \%$ are infected with HPV11. However, majority of these studies targeted specific populations, older age groups, restricted sampling areas, and used only cervical samples (Giuliano et al., 2015; Mbulawa et al., 2017, 2018; Taku et al., 2020). Furthermore, no largescale South African studies or Free State based studies report on HPV11 intratypic variant distribution in patients with RRP, and no current information regarding variants circulating in the Free State exists. Therefore, this study aimed to genetically characterise HPV11 isolates from patients diagnosed with RRP treated at Universitas Academic Hospital.

In a previous unpublished study performed in our laboratory, a novel HPV11 lineage B isolate was identified (Makatsa, 2012). Although not included in this study, this motivated investigation of further isolates for clarification of currently circulating lineages and sub-lineages. Therefore, HPV11 genetic variants were characterised based on the nucleotide sequence of the L1, URR, E5a/b and a segment of E2 genomic regions isolates from patients with RRP in the Free State. According to previous studies, these gene segments encompass sufficient SNPs to discriminate between viral lineages and sublineages, as the concatenated L1-URR-E5a sequence reportedly contains adequate information for differentiation between A1 and A2 sub-lineages. The E5a/b ORF also has a faster evolutionary rate compared to other genes and contains several (sub)lineage specific SNPs. Furthermore, it has been reported that the E2 gene segment can reliably distinguish all known HPV11 variants (Bravo \& Alonso, 2004; Burk et al., 2011; Godínez et al., 2014; Jelen et al., 2016; Maver et al., 2011). A previous study
published by Jelen and colleagues in 2014 suggested that concatenated E5a/b-L1-URR sequences can be used as a surrogate for the phylogenetic clustering of HPV6 variants, a LR-HPV type (Jelen et al., 2014). However, results obtained in the current study uncovered various classification errors using concatenated E5a/b-L1-URR sequence data and does not support the use of these gene segments for HPV11 variant classification.

In Chapter 2, the nucleotide- and amino acid composition of the L1-, E5a/b-, and a segment of the E2 ORF, as well as the nucleotide sequence of the URR were determined using Sanger sequencing. Although nucleotide changes may imply a variant lineage or sub-lineage, these variances do not necessarily impact the expressed protein function. However, DNA alterations in an ORF may lead to gene product alterations if non-synonymous substitutions are present (Lebeuf-Taylor et al., 2019). Few non-synonymous substitutions were reported in the L1 ORF of sequenced isolates, unlike the E5a/b ORF, where nearly all nucleotide substitutions were non-synonymous. The URR is a variable region within the HPV genome capable of accumulating and tolerating many nucleotide mutations as it does not encode proteins. Nevertheless, variations in the URR may still impact virus survival as this region contains crucial binding sites and promoters (Fang et al., 2020; Ribeiro et al., 2018). Amino acid changes may affect many factors such as the effectiveness of HPV11 targeted vaccines, protein folding, cell transformation, the viral life cycle, and disease severity. However, to fully understand the impact if these changes functional analysis of HPV11 genes is required to decipher the effects of mutations on the virus. Consequences of these polymorphisms in the novel lineage B and sub-lineage A3 and A4 warrants further investigation.

Analysis of the results obtained in the current study included constructing a maximum likelihood tree to confirm the differentiation of the HPV variants based on tree topology and to determine the presence of potentially novel intratypic variants. HPV11 concatenated L1-URR and E5a/b-L1-URR resolved into two lineages, namely lineage A and lineage B and the majority of the isolates sequenced (48/67) clustered together with the globally predominant A2 prototype isolates as anticipated. Two lineage B isolates were identified in the current study, VBD34/08 and VBD28/14, with VBD34/08 appearing identical to the previously identified lineage B isolate (GenBank accession number LN833183.1) (Jelen et al., 2016). The concatenated L1-URR, E5a/b-L1-URR sequences, and the E2 segment provided equivalent discriminatory power to distinguish between HPV-lineages; however, tree topology based on E5a/b-L1-URR sequences were incoherent for sub-lineage classification.

Previous studies of intratypic evolution of HPV11 variants rarely dealt with whole-genome sequences. Instead, partial regions of the viral genome such as L1, URR, and an ORF gene were used (de Matos et al., 2013; Danielewski et al., 2013), leading to classification errors, as HPV genome variants are very closely related and only require a $0.5 \%$ to $1 \%$ variance in the complete genome (Burk et al., 2011). Therefore, the complete genome of selected novel HPV11 variants and variants of interest were
determined in Chapter 3 using NGS technology. Analysis of the whole genome ensures accurate differentiation into sub-lineages and detection of any heterogeneity of HPV11 (Burk et al., 2011).

Based on complete sequence data, the majority of nucleotide substitutions and indels were identified in VBD28/14. Non-synonymous substitutions were identified in ORFs of each of the four isolates which may alter protein function (Lebeuf-Taylor et al., 2019), which underlie disease phenotype. Likewise, altered protein function may lead to cancer progression and failure of treatment options (Barra et al., 2019; Brotherton, 2019; Cornall et al., 2013; Gerein et al., 2005; Huebbers et al., 2013). Therefore, future studies regarding viral gene expression are vital. Phylogenetic analysis of complete HPV11 genomes identified VBD28/14 as a lineage B variant, and VBD01/10, VBD15/10 and VBD74/09 as sub-lineage A2 variants with great certainty. However, VBD15/10 and VBD74/09 contained numerous genomic- and amino acid variances compared to the A2 prototype. The slow evolutionary rate of papillomaviruses was reinforced as VBD15/10 and VBD74/09 isolated five months from each other from a single patient were identical (Bernard, 1994; Chen et al., 2009, 2011; van Doorslaer, 2013).

Many inconsistencies were observed between the maximum likelihood trees constructed using the concatenated E5a/b-L1-URR sequences and WGS. Phylogenetic analysis using the concatenated E5a/b-L1-URR sequences revealed that VBD01/10 did not cluster closely with the A2 prototype sequence as was the case using WGS. Also, analysis using the concatenated E5a/b-L1-URR sequences revealed VBD01/10, FN90762.1, and FN907963.1 clustered together with great certainty, However, FN907963.1 is a confirmed sub-lineage A1 isolate according to WGS. Furthermore, phylogenetic analysis revealed VBD15/10, VBD74/09, MK463916.1 and MK463921.1 clustered closely using the concatenated sequences, but this was not the case using complete HPV11 genomes. Discrepancies in the location of clustering of sub-lineage A2 isolates JQ773408.1, JQ773411.1, JQJQ773412.1, MK463914.1, and FR872717.1 was also revealed. Unlike using concatenated sequences, these sequences did not cluster closely with the A2 prototype using the complete HPV11 genome. An important discrepancy observed between using specific genes versus the complete genome was that A1 isolates did not cluster with the corresponding prototype using the concatenated sequences. Therefore, using concatenated E5a/b-L1-URR sequences cannot reliably distinguish between HPV sub-lineages as previously suggested (Jelen et al., 2016).

The 208bp region located at the $3^{\prime}$ end of the E2 ORF and the $5^{\prime}$ end of NCR2 provided more accurate results for HPV11 whole-genome tree reconstruction compared to the concatenated E5a/b-L1-URR sequences. All isolates were classified into appropriate sub-lineages as determined by WGS of the complete HPV11 sequence. This is the second report of a lineage B isolate identified in South Africa (Jelen et al., 2016). Two isolates of interest were identified in the current study. These isolates had numerous amino acid heterogeneities and the consequences of these polymorphisms requires additional investigation. The 208bp E2 segment could reliably classify all isolates in this study, suggesting that
this gene segment contains stable sub-lineage specific SNP's and may serve in sub-lineage identification when complete genome sequences cannot be obtained.

Although this study successfully completed the aim and objectives set out, limitations in the study were identified. The Sanger sequencing approach was used to sequence selected genes of the isolates; however, Sanger sequencing has a significantly lower sensitivity for genomic variant detection compared to NGS which may hinder detection of genomic variants present at low copy numbers (Parker \& Chen, 2017; Sanger et al., 1977). Moreover, research with larger sample sizes will contribute towards clarification of the prevalence, disease association and impact of newly characterised variants on HPV persistence and RRP development. Functional analysis of genes would add valuable information regarding protein interactions and functioning.

This study provides the most comprehensive data on the genomic diversity of HPV11 from patients with RRP in the Free State to date. As within type variants may have an association with disease outcome, it is essential to characterise the current strains circulating amongst the community. Baseline data on circulating HPV variants may significantly contribute to future evolutionary, epidemiological, vaccination, and molecular assay development studies, as well as to studies on the pathogenesis of HPV11 and other Alphapapillomaviruses.

## References

Abreu, AL., Souza, RP., Gimenes, F. \& Consolaro, ME. (2012). A review of methods for detect human papillomavirus infection. Virology Journal, 9, 262. DOI: 10.1186/1743-422X-9-262

Allawi, HT. \& SantaLucia, J. (1997). Thermodynamics and NMR of internal G-T mismatches in DNA. Biochemistry, 36(34), 10581-10594. DOI: 10.1021/bi962590c

Altamura, G., Power, K., Martano, M., Degli Uberti, B., Galiero, G., De Luca, G., Maiolino, P. \& Borzacchiello, G. (2018). Felis catus papillomavirus type-2 E6 binds to E6AP, promotes E6AP/p53 binding and enhances p53 proteasomal degradation. Scientific Reports, 8(1), 17529. DOI: 10.1038/s41598-018-35723-7

Altamura, G., Tommasino, M. \& Borzacchiello, G. (2020). Cutaneous vs. Mucosal Tropism: The papillomavirus paradigm comes to an "and". Frontiers in Microbiology, 11, 588663. DOI: 10.3389/fmicb. 2020.588663

Ambulos, N., Schumaker, L., Mathias, TJ., White, R., Troyer, J., Wells, D. \& Cullen, K. (2016). Nextgeneration sequencing-based HPV genotyping assay validated in formalin-fixed, paraffin-embedded oropharyngeal and cervical cancer specimens. Journal of Biomolecular Techniques, 27(2),46-52. DOI: 10.7171/jbt.16-2702-004

Balaban, M., Moshiri, N., Mai, U., Jia, X. \& Mirarab, S. (2019). TreeCluster: Clustering biological sequences using phylogenetic trees. PLOS One, 14(8), e0221068. DOI: 10.1371/journal.pone. 0221068

Barra, F., Leone Roberti Maggiore, U., Bogani, G., Ditto, A., Signorelli, M., Martinelli, F., Chiappa, V., Lorusso, D., Raspagliesi, F. \& Ferrero, S. (2019). New prophylactics human papilloma virus (HPV) vaccines against cervical cancer. Journal of Obstetrics and Gynaecology, 39(1), 1-10. DOI: 10.1080/01443615.2018.1493441

Bedell, MA., Hudson, JB., Golub, TR., Turyk, ME., Hosken, M., Wilbanks, GD. \& Laimins, LA. (1991). Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation. Journal of Virology, 65(5), 2254-2260. DOI: 10.1128/JVI.65.5.2254-2260.1991

Benedict, JJ. \& Derkay, CS. (2021). Recurrent respiratory papillomatosis: A 2020 perspective. Laryngoscope Investigative Otolaryngology, 6(2), 340-345. DOI: 10.1002/lio2.545

Bernard, HU. (1994). Coevolution of papillomaviruses with human populations. Trends in Microbiology, 2(4), 140-3. DOI: 10.1016/0966-842x(94)90602-5

Bernard, HU., Burk, RD., Chen, Z., van Doorslaer, K., zur Hausen, H. \& de Villiers, EM. (2010). Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology, 401(1), 70-79. DOI: 10.1016/j. virol.2010.02.002

Bernard, HU., Calleja-Macias IE. \& Dunn, ST. (2006). Genome variation of human papillomavirus types: Phylogenetic and medical implications. International Journal of Cancer, 118(5), 1071-1076. DOI: 10.1002/ijc. 21655

Bonagura, VR., Du, Z., Ashouri, E., Luo, L., Hatam, LJ., DeVoti, JA., Rosenthal, DW., Steinberg, BM., Abramson, AL., Gjertson, DW., Reed, EF. \& Rajalingam, R. (2010). Activating killer cell immunoglobulin-like receptors 3DS1 and 2DS1 protect against developing the severe form of recurrent respiratory papillomatosis. Human Immunology, 71(2), 212-219. DOI: 10.1016/j.humimm.2009.10.009

Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W. \& zur Hausen, H. (1984). A new type of papillomavirus DNA, its presence in genital cancer biopsies and in cell lines derived from cervical cancer. The EMBO Journal, 3(5), 1151-1157. DOI: 10.1002/j.1460-2075.1984.tb01944.x

Bravo, IG. \& Alonso, A. (2004). Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth. Journal of Virology, 78(24), 13613-13626. DOI: 10.1128/JVI.78.24.13613-13626.2004

Brink, AA., Snijders, PJ. \& Meijer, CJ. (2007). HPV detection methods. Disease Markers, 23(4), 273281. DOI: 10.1155/2007/147429

Brotherton, JM. (2019). Impact of HPV vaccination: Achievements and future challenges. Papillomavirus Research, 7, 138-140. DOI: 10.1016/j.pvr.2019.04.004

Buchinsky, FJ., Ruszkay, N., Valentino, W., Derkay, CS., McClay, JE., Bastian, RW., Myer, CM., Lollar, KW. \& Guris, D. (2020). In RRP, serologic response to HPV is frequently absent and slow to develop. PLOS One, 15(3), e0230106. DOI: 10.1371/journal.pone. 0230106

Buck, CB., Day, PM. \& Trus, BL. (2013). The papillomavirus major capsid protein L1. Virology, 445(1-2), 169-174. DOI: 10.1016/j.virol.2013.05.038

Burd, EM. (2016). Human papillomavirus laboratory testing: The changing paradigm. Clinical Microbiology Reviews, 29(2), 291-319. DOI: 10.1128/cmr.00013-15

Burk, RD., Chen, Z., Harari, A., Smith, B. C., Kocjan, B. J., Maver, P. J. \& Poljak, M. (2011). Classification and nomenclature system for human Alphapapillomavirus variants: General features, nucleotide landmarks and assignment of HPV6 and HPV11 isolates to variant lineages. Acta
dermatovenerologica Alpina, Pannonica, et Adriatica, 20(3), 113-123. Retrieved from https://www.acta-apa.org/

Burk, RD., Harari, A. \& Chen, Z. (2013). Human papillomavirus genome variants. Virology, 445(1-2), 232-243. DOI: 10.1016/j.virol.2013.07.018

Bzhalava, D., Eklund, C. \& Dillner, J. (2015). International standardization and classification of human papillomavirus types. Virology, 476, 341-344. DOI: 10.1016/j.virol.2014.12.028

Chabeda, A., Yanez, R., Lamprecht, R., Meyers, AE., Rybicki, EP. \& Hitzeroth, II. (2018). Therapeutic vaccines for high-risk HPV-associated diseases. Papillomavirus Research, 5, 46-58. DOI: 10.1016/j.pvr.2017.12.006

Chagas, BS., Batista, MV., Guimarães, V., Balbino, VQ., Crovella, S. \& Freitas, AC. (2011). New variants of E6 and E7 oncogenes of human papillomavirus type 31 identified in Northeastern Brazil. Gynecologic Oncology, 123(2), 284-288. DOI: 10.1016/j.ygyno.2011.07.008

Charleston, M. (2013). Phylogeny, in Maloy, S., Hughes, K (eds) Brenner's Encyclopaedia of Genetics. $2^{\text {nd }}$ edition. Academic Press, 324,325. DOI: 10.1016/B978-0-12-374984-0.01160-8

Chen, Z., DeSalle, R., Schiffman, M., Herrero, R. \& Burk, RD. (2009). Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97. Journal of Virology, 83(3), 1443-1455. DOI: 10.1128/jvi.02068-08

Chen, Z., Schiffman, M., Herrero, R., Desalle, R., Anastos, K., Segondy, M., Sahasrabuddhe, VV., Gravitt, P. E., Hsing, A. W. \& Burk, R. D. (2011). Evolution and taxonomic classification of human papillomavirus 16 (HPV16)-related variant genomes: HPV31, HPV33, HPV35, HPV52, HPV58 and HPV67. PLOS One, 6(5), e20183. DOI: 10.1371/journal.pone. 0020183

Chowdhury, B., Garai, A. \& Garai, G. (2017). An optimized approach for annotation of large eukaryotic genomic sequences using genetic algorithm. BMC bioinformatics, 18(1), 460. DOI: 10.1186/s12859-017-1874-7

Combrinck, CE., Seedat, RY., Randall, C., Roodt, Y. \& Burt, FJ. (2012). Novel HPV-6 variants of human papillomavirus causing recurrent respiratory papillomatosis in southern Africa. Epidemiology and Infection, 140(6), 1095-1101. DOI: 10.1017/s0950268811001580

Conway, MJ. \& Meyers, C. (2009). Replication and assembly of human papillomaviruses. Journal of Dental Research, 88(4), 307-317. DOI: 10.1177/0022034509333446

Cornall, AM., Roberts, JM., Garland, SM., Hillman, RJ., Grulich, AE. \& Tabrizi, SN. (2013). Anal and perianal squamous carcinomas and high-grade intraepithelial lesions exclusively associated with "lowrisk" HPV genotypes 6 and 11. International Journal of Cancer, 133(9), 2253-2258. DOI: 10.1002/ijc. 28228

Cornelissen, MT., Bots, T., Briët, MA., Jebbink, M. F., Struyk, AP., van den Tweel, JG., Greer, CE., Smits, HL. \& ter Schegget, J. (1992). Detection of human papillomavirus types by the polymerase chain reaction and the differentiation between high-risk and low-risk cervical lesions. Virchows Archiv, 62(3), 167-171. DOI: 10.1007/bf02899679

Crawford, LV. (1965). A study of human papilloma virus DNA. Journal of Molecular Biology, 13(2), 362-372. DOI: $10.1016 / \mathrm{s} 0022-2836(65) 80103-6$

D'Abramo, CM. \& Archambault, J. (2011). Small molecule inhibitors of human papillomavirus protein - protein interactions. The Open Virology Journal, 5, 80-95. DOI: 10.2174/1874357901105010080

Dadar, M., Chakraborty, S., Dhama, K., Prasad, M., Khandia, R., Hassan, S., Munjal, A., Tiwari, R., Karthik, K., Kumar, D., Iqbal, H. \& Chaicumpa, W. (2018). Advances in designing and developing vaccines, drugs and therapeutic approaches to counter human papillomavirus. Frontiers in Immunology, 9, 2478. DOI: $10.3389 /$ fimmu. 2018.02478

Danielewski, JA., Garland, SM., McCloskey, J., Hillman, RJ. \& Tabrizi, SN. (2013). Human papillomavirus type 6 and 11 genetic variants found in 71 oral and anogenital epithelial samples from Australia. PLOS One, 8(5), e63892. DOI: 10.1371/journal.pone. 0063892

Dartmann, K., Schwarz, E., Gissmann, L. \& zur Hausen, H. (1986). The nucleotide sequence and genome organization of human papilloma virus type 11. Virology, 151(1), 124-130. DOI: 10.1016/0042-6822(86)90110-8
de Matos, RP., Sichero, L., Mansur, IM., do Bonfim, CM., Bittar, C., Nogueira, RL., Küpper, DS., Valera, FC., Nogueira, ML., Villa, LL., Calmon, MF. \& Rahal, P. (2013). Nucleotide and phylogenetic analysis of human papillomavirus types 6 and 11 isolated from recurrent respiratory papillomatosis in Brazil. Infection, Genetics and Evolution, 16, 282-289. DOI: 10.1016/j.meegid.2012.12.033
de Roda Husman, AM., Walboomers, JM., van den Brule, AJ., Meijer, CJ. \& Snijders, PJ. (1995). The use of general primers GP5 and GP6 elongated at their 3' ends with adjacent highly conserved sequences improves human papillomavirus detection by PCR. Journal of General Virology, 76(4), 1057-1062. DOI: 10.1099/0022-1317-76-4-1057
de Villiers, EM. (2013). Cross-roads in the classification of papillomaviruses. Virology, 445(1-2), 2-10. DOI: 10.1016/j.virol.2013.04.023
de Villiers, EM., Fauquet, C., Broker, TR., Bernard, HU. \& zur Hausen, H. (2004). Classification of papillomaviruses. Virology, 324(1), 17-27. DOI: 10.1016/j.virol.2010.02.002

Delany-Moretlwe, S., Kelley, KF., James, S., Scorgie, F., Subedar, H., Dlamini, NR., Pillay, Y., Naidoo, N., Chikandiwa, A. \& Rees, H. (2018). Human papillomavirus vaccine introduction in South Africa: Implementation lessons from an evaluation of the national school-based vaccination campaign. Global Health: Science and Practice, 6(3), 425-438. DOI: 10.9745/ghsp-d-18-00090

DiLorenzo, TP. \& Steinberg, BM. (1995). Differential regulation of human papillomavirus type 6 and 11 early promoters in cultured cells derived from laryngeal papillomas. Journal of Virology, 69(11), 6865-6872. DOI: 10.1128/jvi.69.11.6865-6872.1995

Dixit, R., Bhavsar, C. \& Marfatia, YS. (2011). Laboratory diagnosis of human papillomavirus infection in female genital tract. Indian Journal of Sexually Transmitted Diseases and AIDS, 32(1), 50. DOI: 10.4103/2589-0557.81257

Dollard, SC., Wilson, JL., Demeter, LM., Bonnez, W., Reichman, RC., Broker, TR. \& Chow, LT. (1992). Production of human papillomavirus and modulation of the infectious program in epithelial raft cultures. Genes and Development, 6(7), 1131-1142. DOI: 10.1101/gad.6.7.1131

Doorbar, J., Quint, W., Banks, L., Bravo, IG., Stoler, M., Broker, TR. \& Stanley, MA. (2012). The biology and life-cycle of human papillomaviruses. Vaccine, 30(5), F55-F70. DOI: 10.1016/j.vaccine.2012.06.083

Dürst, M, Gissmann, L., Ikenberg, H \& zur Hausen, H. (1983). A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. Proceedings of the National Academy of Sciences of the United States of America, 3812-3815. DOI: 10.1073/pnas.80.12.3812

Egawa, N. \& Doorbar, J. (2017). The low-risk papillomaviruses. Virus Research, 231, 119-127. DOI: 10.1016/j.virusres.2016.12.017

Egawa, N., Egawa, K., Griffin, H. \& Doorbar, J. (2015). Human papillomaviruses; epithelial tropisms, and the development of neoplasia. Viruses, 7(7), 3863-3890. DOI: 10.3390/v7072802

El Achkar, VNR., Costa, V., Sichero, L., Ribeiro-Silva, A., Carlos, R., Villa, LL. \& Kaminagakura, E. (2020). Inflammatory cells and HPV type related to aggressiveness in laryngeal papillomatosis. Oral

Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 30(3) e275. DOI: 10.1016/j.0000.2020.04.752

Fang, L., Lin, X., Yang, Y., Song, Z., Ding, X., Tan, L. \& Gao, P. (2020). Genetic variability, phylogeny and functional implication of the long control region in human papillomavirus type 16,18 and 58 in Chengdu, China. Virology Journal, 17(1), 1-15. DOI: 10.1186/s12985-020-01349-3

Fauquet, CM., Mayo, MA., Maniloff, J., Desselberger, U. \& Ball, LA. (2005). VIIIth report of the International Committee on Taxonomy of Viruses, in CM, Fauquet; MA, Mayo; J, Maniloff; U, Desselberger \& LA, Ball (eds). Virus Taxonomy. Academic Press, (83), 988-992

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution, 17(6), 368-376. DOI: 10.1007/bf01734359

Flores-Díaz, E., Sereday, KA., Ferreira, S., Sirak, B., Sobrinho, JS., Baggio, ML., Galan, L., Silva, RC., Lazcano-Ponce, E., Giuliano, AR., Villa, LL., and Sichero, L. (2017). HPV-6 molecular variants association with the development of genital warts in men: The HIM Study, The Journal of Infectious Diseases, 215(4), 559-565. DOI: 10.1093/infdis/jiw600

Fusconi, M., Grasso, M., Greco, A., Gallo, A., Campo, F., Remacle, M., Turchetta, R., Pagliuca, G. \& DE Vincentiis, M. (2014). Recurrent respiratory papillomatosis by HPV: Review of the literature and update on the use of cidofovir. Acta otorhinolaryngologica Italica, 34(6), 375-381. Retrieved from https://www.actaitalica.it/index

Gáll, T., Kis, A., Tatár, TZ., Kardos, G., Gergely, L. \& Szarka, K. (2013). Genomic differences in the background of different severity in juvenile-onset respiratory papillomatoses associated with human papillomavirus type 11. Medical Microbiology and Immunology. 202, 353-363. DOI: 10.1007/s00430-013-0297-y

Gaylis, B. \& Hayden, RE. (1991). Recurrent respiratory papillomatosis: Progression to invasion and malignancy. American Journal of Otolaryngology, 12(2), 104-112. DOI: 10.1016/0196-0709(91)90045-h

Gazia, F., Galletti, B., Freni, F., Bruno, R., Sireci, F., Galletti, C., Meduri, F. \& Galletti, F. (2020). Use of intralesional cidofovir in the recurrent respiratory papillomatosis: A review of the literature. European Review for Medical and Pharmacological Sciences, 24(2), 956-962. DOI: 10.26355/eurrev_202001_20081

Georgescu, SR., Mitran, CI., Mitran, MI., Caruntu, C., Sarbu, MI., Matei, C., Nicolae, I., Tocut, SM., Popa, MI. \& Tampa, M. (2018). New insights in the pathogenesis of HPV infection and the associated
carcinogenic processes: The role of chronic inflammation and oxidative stress. Journal of Immunology Research, 2018, 1-10. DOI: 10.1155/2018/5315816

Gerein, V., Rastorguev, E., Gerein, J., Draf, W. \& Schirren, J. (2005). Incidence, age at onset, and potential reasons of malignant transformation in recurrent respiratory papillomatosis patients: 20 years' experience. Otolaryngology—Head and Neck Surgery, 132(3), 392-394. DOI: 10.1016/j.otohns.2004.09.035

Gissmann, L., Diehl, V., Schultz-Coulon, HJ \& zur Hausen, H. (1982). Molecular cloning and characterization of human papilloma virus DNA derived from laryngeal papilloma. Journal of Virology, 44, 393-400. DOI: 10.1128/jvi.44.1.393-400.1982

Gissmann, L., Pfister, H. \& zur Hausen, H. (1977). Human papilloma viruses (HPV): Characterization of four different isolates. Virology, 76(2), 569-580. DOI: 10.1016/0042-6822(77)90239-2

Gissmann, L., Wolnik, L., Ikenberg, H., Koldovsky, U., Schnürch, HG. \& zur Hausen, H. (1983). Human papillomavirus types 6 and 11 DNA sequences in genital and laryngeal papillomas and in some cervical cancers. Proceedings of the National Academies of Science, 80, 560-563. DOI: 10.1073/pnas.80.2.560

Giuliano, AR., Botha, MH., Zeier, M., Abrahamsen, ME., Glashoff, RH., van der Laan, LE., Papenfuss, M., Engelbrecht, S., Schim van der Loeff, MF., Sudenga, SL., Torres, BN., Kipping, S. \& Taylor, D. (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, 68(2), 227-35. DOI: 10.1097/qai. 0000000000000425

Godínez, JM., Nicolás-Párraga, S., Pimenoff, VN., Mengual-Chuliá, B., Muñoz, N., Bosch, FX., Sánchez, GI., McCloskey, J. \& Bravo IG. (2014). Phylogenetically related, clinically different: Human papillomaviruses 6 and 11 variants distribution in genital warts and in laryngeal papillomatosis. Clinical Microbiology and Infection, 20(6), O406-O413. DOI: 10.1111/1469-0691.12420

Graham, SV. \& Faizo, A. (2017). Control of human papillomavirus gene expression by alternative splicing. Virus Research, 231, 83-95. DOI: 10.1016/j.virusres.2016.11.016

Grasso, M., Remacle, M., Bachy, V., van der Vorst, S. \& Lawson, G. (2014). Use of cidofovir in HPV patients with recurrent respiratory papillomatosis. European Archives of Oto-Rhino-Laryngology, 271(11), 2983-2990. DOI: 10.1007/s00405-014-3055-x

Graupp, M., Gugatschka, M., Kiesler, K., Reckenzaun, E., Hammer, G. \& Friedrich, G. (2013). Experience of 11 years use of cidofovir in recurrent respiratory papillomatosis. European Archives of Oto-Rhino-Laryngology, 270(2), 641-646. DOI: 10.1007/s00405-012-2221-2

Hahn, HS., Kee, MK., Kim, HJ., Kim, MY., Kang, YS., Park, JS., Kim, TJ. (2013). Distribution of maternal and infant human papillomavirus: Risk factors associated with vertical transmission. European Journal of Obstetrics \& Gynecology and Reproductive Biology, 169(2), 202-206. DOI: 10.1016/j.ejogrb.2013.02.024

Halavaty, KK., Regan, J., Mehta, K. \& Laimins, L. (2014). Human papillomavirus E5 oncoproteins bind the A4 endoplasmic reticulum protein to regulate proliferative ability upon differentiation. Virology. 452, 223-30. DOI: 10.1016/j.virol.2014.01.013

Hall, BG. (2013). Building phylogenetic trees from molecular data with MEGA. Molecular Biology and Evolution, 30(5), 1229-1235. DOI: 10.1093/molbev/mst012

Harari, A., Chen, Z. \& Burk, RD. (2014). Human papillomavirus genomics: Past, present and future. Current Problems in Dermatology, 45, 1-18. DOI: 10.1159/000355952

Hariri, S., Schuler, MS., Naleway, AL., Daley, MF., Weinmann, S., Crane, B., Newcomer, SR., Tolsma, D. \& Markowitz, LE. (2018). Human papillomavirus vaccine effectiveness against incident genital warts among female health-plan enrollees, United States. American Journal of Epidemiology, 187(2), 298-305. DOI: 10.1093/aje/kwx253

Heinzel, PA., Chan, SY., Ho, L., O’Connor, M., Balaram, P., Campo, MS., Fujinaga, K., Kiviat, N., Kuypers, J., Pfister, H., Steinberg, BM., Tay, SK., Villa, LL. \& Bernard, HU. (1995). Variation of human papillomavirus type 6 (HPV-6) and HPV-11 genomes sampled throughout the world. Journal of Clinical Microbiology, 33, 1746-1754. DOI: 10.1128/jcm.33.7.1746-1754.1995

Hemmat, N. \& Baghi, HB. (2018). Human papillomavirus E5 protein, the undercover culprit of tumorigenesis. Infectious Agents and Cancer, 13, 31. DOI: 10.1186/s13027-018-0208-3

Hirose, Y., Onuki, M., Tenjimbayashi, Y., Mori, S., Ishii, Y., Takeuchi, T., Tasaka, N., Satoh, T., Morisada, T., Iwata, T., Miyamoto, S., Matsumoto, K., Sekizawa, A. \& Kukimoto, I. (2018). Withinhost variations of human papillomavirus reveal APOBEC signature mutagenesis in the viral genome. Journal of Virology, 92(12), e00017-18. DOI: 10.1128/jvi.00017-18

Hoesli, RC., Wingo, ML., Richardson, BE. \& Bastian, RW. (2020). Identification of 11 different HPV subtypes in adult patients with recurrent respiratory papillomatosis. Otolaryngology-Head and Neck Surgery, 163(4), 785-790. DOI: 10.1177/0194599820931817

Huebbers, CU., Preuss, SF., Kolligs, J., Vent, J., Stenner, M., Wieland, U., Silling, S., Drebber, U., Speel, EJM. \& Klussmann, JP. (2013). Integration of HPV6 \& downregulation of AKR1C3 expression mark malignant transformation in a patient with juvenile-onset laryngeal papillomatosis. PLOS One, 8(2), e57207. DOI: 10.1371/journal.pone. 0057207

Intakorn, P. \& Sonsuwan, N. (2014). Human papillomatosis genotyping and severity in patients with recurrent respiratory papillomatosis. Journal of the Medical Association of Thailand, 97 (6), S136S141

International Committee on Taxonomy of Viruses. (2021). Taxonomy, viewed on 22 July 2021, [https://talk.ictvonline.org/taxonomy/](https://talk.ictvonline.org/taxonomy/)

Ivancic, R., Iqbal, H., deSilva, B., Pan, Q. \& Matrka, L. (2018). Current and future management of recurrent respiratory papillomatosis. Laryngoscope Investigative Otolaryngology, 3(1), 22-34. DOI: 10.1002/lio2.132

Ivancic, R., Iqbal, H., deSilva, B., Pan, Q. \& Matrka, L. (2020). Immunological tolerance of low-risk HPV in recurrent respiratory papillomatosis. Clinical and Experimental Immunology, 199(2), 131-142. DOI: 10.1111/cei. 13387

Jelen, MM., Chen, Z., Kocjan, BJ., Burt, FJ., Chan, PK., Chouhy, D., Combrinck, CE., Coutlée, F., Estrade, C., Ferenczy, A., Fiander, A., Franco, EL., Garland, SM., Giri, AA., González, JV., Gröning, A., Heidrich, K., Hibbitts, S., Hošnjak, L., Luk, TN., Marinic, K., Matsukura, T., Neumann, A., Oštrbenk, A., Picconi, MA., Richardson, H., Sagadin, M., Sahli, R., Seedat, RY., Seme, K., Severini, A., Sinchi, JL., Smahelova, J., Tabrizi, SN., Tachezy, R., Tohme, S., Uloza, V., Vitkauskiene, A., Wong, YW., Zidovec Lepej, S., Burk, RD. \& Poljak, M. (2014). Global genomic diversity of human papillomavirus 6 based on 724 isolates and 190 complete genome sequences. Journal of Virology, 88, 7307-7316. DOI: 10.1128/jvi.00621-14

Jelen, MM., Chen, Z., Kocjan, BJ., Hošnjak, L., Burt, FJ., Chan, PK., Chouhy, D., Combrinck, CE., Estrade, C., Fiander, A., Garland, SM., Giri, AA., González, JV., Gröning, A., Hibbitts, S., Luk, TNM., Marinic, K., Matsukura, T., Neumann, A., Oštrbenk, A., Picconi, MA., Sagadin, M., Sahli, R., Seedat, RY., Seme, K., Severini, A., Sinchi, JL., Smahelova, J., Tabrizi, SN., Tachezy, R., Tohme Faybush, S., Uloza, V., Uloziene, I., Wong, YW., Židovec Lepej, S., Burk, RD. \& Poljak, M. (2016). Global genomic diversity of human papillomavirus 11 based on 433 isolates and 78 complete genome sequences. Journal of Virology, 90(11), 5503-5513. DOI: 10.1128/jvi.03149-15

Kardani, K. \& Bolhassani, A. (2018a). HPV proteins and their functions, in A, Bolhassani. HPV Infections: Diagnosis, Prevention and Treatment. Bentham Science Publishers, 8-20. DOI: 10.2174/97816810861701180101

Kardani, K. \& Bolhassani, A. (2018b). Types of benign or malignant diseases associated with hpv infection, in A, Bolhassani. HPV Infections: Diagnosis, Prevention and Treatment. Bentham Science Publishers, 30-40. DOI: 10.2174/97816810861701180101

Kardani, K., Naderi, N. \& Bolhassani, A. (2018). The life cycle and transmission of hpv types, in N, Naderi; K, Kardani \& Bolhassani. HPV Infections: Diagnosis, Prevention, and Treatment. Bentham Science Publishers, 44-52. DOI: 10.2174/97816810861701180101

Klingelhutz, AJ. \& Roman, A. (2012). Cellular transformation by human papillomaviruses: Lessons learned by comparing high-and low-risk viruses. Virology, 424(2), 77-98. DOI: 10.1016/j.virol. 2011.12 .018

Klug, A. \& Finch, JT. (1965). Structure of viruses of the papilloma-polyoma type: I. Human wart virus. Journal of Molecular Biology, 11(2), 403-423. DOI: 10.1016/s0022-2836(65)80066-3

Kocjan, BJ., Bzhalava, D., Forslund, O., Dillner, J. \& Poljak, M. (2015). Molecular methods for identification and characterization of novel papillomaviruses. Clinical Microbiology and Infection, 21(9), 808-816. DOI: 10.1016/j.cmi.2015.05.011

Kuhner, MK. \& Felsenstein, J. (1994). A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. Molecular Biology and Evolution, 11(3), 459-68. DOI: 10.1093/oxfordjournals.molbev.a040126

Kumar, S., Stecher, G., Li, M., Knyaz, C. \& Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution, 35(6), 1547-1549. DOI: 10.1093/molbev/msy096

Larkin, MA., Blackshields, G., Brown, NP., Chenna, R., McGettigan, PA., McWilliam, H., Valentin, F., Wallace, IM., Wilm, A., Lopez, R., Thompson, JD., Gibson, TJ. \& Higgins, DG. (2007). Clustal W and Clustal X version 2.0. Bioinformatics, 23, 2947-2948. DOI: 10.1093/bioinformatics/btm404

Lebeuf-Taylor, E., McCloskey, N., Bailey, SF., Hinz, A. \& Kassen, R. (2019). The distribution of fitness effects among synonymous mutations in a gene under directional selection. eLife, 8, e45952. DOI: 10.7554/eLife. 45952

Liu, X., Roberts, J., Dakic, A., Zhang, Y. \& Schlegel, R. (2008). HPV E7 contributes to the telomerase activity of immortalized and tumorigenic cells and augments E6-induced hTERT promoter function. Virology, 375(2), 611-623. DOI: 10.1016/j.virol.2008.02.025

Longworth, MS. \& Laimins, LA. (2004). Pathogenesis of human papillomaviruses in differentiating epithelia. Microbiology and Molecular Biology Reviews, 68(2), 362-372. DOI: 10.1128/mmbr.68.2.362-372.2004

Madison, KC. (2003). Barrier function of the skin: "la raison d'etre" of the epidermis. Journal of Investigative Dermatology, 121(2), 231-241. DOI: 10.1046/j.1523-1747.2003.12359.x

Makatsa, S. (2012). Detection of HPV-11 variants associated with recurrent laryngeal papillomatosis based on LCR-E6 sequence data, Southern Africa. BMedSc Hons mini dissertation. University of the Free State, Bloemfontein

Mattoon, D., Gupta, K., Doyon, J., Loll, PJ. \& DiMaio, D. (2001). Identification of the transmembrane dimer interface of the bovine papillomavirus E5 protein. Oncogene, 20(29) 3824-3834. DOI: 10.1038/sj.onc. 1204523

Maver, PJ., Kocjan, BJ., Seme, K., Potočnik, M., Gale, N. \& Poljak, M. (2011). Prevaccination genomic diversity of human papillomavirus genotype 11: A study on 63 clinical isolates and 10 fulllength genome sequences. Journal of Medical Virology, 83, 461-470. DOI: 10.1002/jmv. 21994.

Mbulawa, Z., Kularatne, R., Kufa-Chakezha, T. \& Williamson, AL. (2017). Sentinel surveillance of human papillomavirus genotypes among patients attending public healthcare facilities in South Africa. 16(2), 1-8. Retrieved from http://nicd.ac.za

Mbulawa, ZZ., Van Schalkwyk, C., Hu, NC., Meiring, TL., Barnabas, S., Dabee, S., Heather Jaspan, H., Kriek, Z. Jaumdally, SZ., Muller, E., Bekker, LG., Lewis, DA., Dietrich, J., Gray, G., Passmore, JAS. \& Williamson, AL. (2018). High human papillomavirus (HPV) prevalence in South African adolescents and young women encourages expanded HPV vaccination campaigns. PLOS One, 13(1), e0190166. DOI: 10.1371/journal.pone. 0190166

McBride, AA. (2017). Mechanisms and strategies of papillomavirus replication. Biological Chemistry, 398(8), 919-927. DOI: 10.1515/hsz-2017-0113

Miguel Rocha, M. \& Ferreira, PG. (2018). Phylogenetic Analysis, in M, Miguel Rocha \& PG, Ferreira. Bioinformatics Algorithms. Academic Press, 199-220

Murahwa, AT., Nindo, F., Onywera, H., Meiring, TL., Martin, DP. \& Williamson, AL. (2019). Evolutionary dynamics of ten novel Gamma-PVs: Insights from phylogenetic incongruence,
recombination and phylodynamic analyses. BMC Genomics, 20(1), 368. DOI: 10.1186/s12864-019-5735-9

Novakovic, D., Cheng, ATL., Zurynski, Y., Booy, R., Walker, PJ., Berkowitz, R., Harrison, H., Black, R., Perry, C., Vijayasekaran, S., Wabnitz, D., Burns, H., Tabrizi, SN., Garland, SM., Elliott, E. \& Brotherton, JML. (2018). A prospective study of the incidence of juvenile-onset recurrent respiratory papillomatosis after implementation of a national HPV vaccination program. The Journal of Infectious Diseases, 217(2), 208-212. DOI: 10.1093/infdis/jix498

Okonechnikov, K., Golosova, O., Fursov, M. \& the UGENE team. (2012). Unipro UGENE: A unified bioinformatics toolkit. Bioinformatics, 28, 1166-1167. DOI: 10.1093/bioinformatics/bts091

Omland, T., Akre, H., Lie, KA., Jebsen, P., Sandvik, L. \& Brøndbo, K. (2014). Risk factors for aggressive recurrent respiratory papillomatosis in adults and juveniles. PLOS One, 9(11), el13584. DOI: 10.1371/journal.pone. 0113584

Pal, A. \& Kundu, R. (2020). Human papillomavirus E6 and E7: The cervical cancer hallmarks and targets for therapy. Frontiers in Microbiology, 10, 3116. DOI: https://doi.org/10.3389/fmicb.2019.03116

Parker, J. \& Chen, J. (2017). Application of next generation sequencing for the detection of human viral pathogens in clinical specimens. Journal of Clinical Virology, 86, 20-26. DOI: 10.1016/j.jcv.2016.11.010

Pastrana, DV., Buck, CB., Pang, YY., Thompson, CD., Castle, PE., FitzGerald, PC., Krüger Kjaer, S., Lowy, DR. \& Schiller, JT. (2004). Reactivity of human sera in a sensitive, high-throughput pseudovirus-based papillomavirus neutralization assay for HPV16 and HPV18. Virology, 321(2), 205216. DOI: 10.1016/j.virol.2003.12.027

Peñaloza-Plascencia, M., Montoya-Fuentes, H., Flores-Martínez, SE., Fierro-Velasco, FJ., PeñalozaGonzález, JM. \& Sánchez-Corona, J. (2000). Molecular identification of 7 human papillomavirus types in recurrent respiratory papillomatosis. Archives of Otolaryngology-Head and Neck Surgery, 126(9), 1119-1123. DOI: 10.1001/archotol.126.9.1119

Pinidis, P., Tsikouras, P., Iatrakis, G., Zervoudis, S., Koukouli, Z., Bothou, A., Galazios, G. \& Vladareanu, S. (2016). Human papilloma virus' life cycle and carcinogenesis. Maedica, 11(1), 48-54. Retrieved from https://www.maedica.ro/

Reidy, PM., Dedo, HH., Rabah, R., Field, JB., Mathog, RH., Gregoire, L. \& Lancaster, WD. (2004). Integration of human papillomavirus type 11 in recurrent respiratory papilloma-associated cancer. The Laryngoscope, 114(11), 1906-1909. DOI: 10.1097/01.mlg.0000147918.81733.49

Ribeiro, AL., Caodaglio, AS. \& Sichero, L. (2018). Regulation of HPV transcription. Clinics, 73(suppl 1), e486s. DOI: $10.6061 /$ clinics/2018/e486s

Rintala, MA., Grénman, SE., Puranen, MH., Isolauri, E., Ekblad, U., Kero, PO. \& Syrjänen, SM. (2005). Transmission of high-risk human papillomavirus (HPV) between parents and infant: A prospective study of HPV in families in Finland. Journal of Clinical Microbiology, 43(1), 376-381. DOI: 10.1128/jcm.43.1.376-381.2005

Rodier, C., Lapointe, A., Coutlée, F., Mayrand, M. H., Dal Soglio, D., Roger, M. \& Trottier, H. (2013). Juvenile respiratory papillomatosis: Risk factors for severity. Journal of Medical Virology, 85(8), 1447-1458. DOI: 10.1002/jmv. 23615

Rodríguez-Álvarez, MI., Gómez-Urquiza, JL., Husein-El Ahmed, H., Albendín-García, L., GómezSalgado, J. \& Cañadas-De la Fuente, GA. (2018). Prevalence and risk factors of human papillomavirus in male patients: A systematic review and meta-analysis. International Journal of Environmental Research and Public Health, 15(10), 2210. DOI: 10.3390/ijerph15102210

Rosenberg, T., Philipsen, BB., Mehlum, CS., Dyrvig, AK., Wehberg, S., Chirilǎ, M. \& Godballe, C. (2019). Therapeutic use of the human papillomavirus vaccine on recurrent respiratory papillomatosis: A systematic review and meta-analysis. The Journal of Infectious Diseases, 219(7), 1016-1025. DOI: 10.1093/infdis/jiy616

Rous, P. \& Beard, JW. (1935). The progression to carcinoma of virus-induced rabbit papillomas (Shope). The Journal of Experimental Medicine, 62(4), 523-48. DOI: 10.1084/jem.62.4.523

Rusinko, J. \& McPartlon, M. (2017). Species tree estimation using Neighbor Joining. Journal of Theoretical Biology, 414, 5-7. DOI: 10.1016/j.jtbi.2016.11.005

Sabeena, S., Bhat, P., Kamath, V. \& Arunkumar, G. (2017). Possible non-sexual modes of transmission of human papilloma virus. Journal of Obstetrics and Gynaecology Research, 43(3), 429-435. DOI: 10.1111/jog. 13248

Saitou, N. \& Nei, M. (1987). The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4(4), 406-425. DOI: 10.1093/oxfordjournals.molbev.a040454

Sanger, F., Nicklen, S. \& Coulson, AR. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America, 74(12), 5463-5467. DOI: 10.1073/pnas.74.12.5463

Seedat, RY. \& Schall, R. (2018). Age of diagnosis, incidence and prevalence of recurrent respiratory papillomatosis—A South African perspective. Clinical Otolaryngology, 43, 533-537. DOI: 10.1111/coa. 13016

Seedat, RY. (2020). Juvenile-onset recurrent respiratory papillomatosis diagnosis and management - A developing country review. Pediatric Health, Medicine and Therapeutics, 11, 39-46. DOI: 10.2147/phmt.s200186

Seedat, RY., Combrinck, CE. \& Burt, FJ. (2013). HPV associated with recurrent respiratory papillomatosis. Future Virology, 8(5), 477-492. DOI: 10.2217/fvl.13.31

Serdoz, S., Egri-Nagy, A., Sumner, J., Holland, BR., Jarvis, PD., Tanaka, MM. \& Francis, AR. (2017). Maximum likelihood estimates of pairwise rearrangement distances. Journal of Theoretical Biology, 423, 31-40. DOI: 10.1016/j.jtbi.2017.04.015

Shope, RE. \& Hurst, EW. (1933). Infectious papillomatosis of rabbits: With a note on the histopathology. The Journal of Experimental Medicine, 58(5), 607-24. DOI: 10.1084/jem.58.5.607

Siqueira, JD., Alves, BM., Prellwitz, IM., Furtado, C., Meyrelles, ÂR., Machado, ES., Seuánez, HN., Soares, MA. \& Soares, EA. (2016). Identification of novel human papillomavirus lineages and sublineages in HIV/HPV-coinfected pregnant women by next-generation sequencing. Virology, 493, 202-208. DOI: 10.1016/j.virol.2016.03.027

Smith, EM., Parker, MA., Rubenstein, LM., Haugen, TH., Hamsikova, E. \& Turek, LP. (2010). Evidence for vertical transmission of HPV from mothers to infants. Infectious Diseases in Obstetrics and Gynecology, 2010, 326369. DOI: 10.1155/2010/326369

So, KA., Lee, IH., Kim, TJ. \& Lee, KH. (2019). Risk factors of persistent HPV infection after treatment for high-grade squamous intraepithelial lesion. Archives of Gynecology and Obstetrics, 299, 223-227. DOI: 10.1007/s00404-018-4936-9

Sridhar, S., To, KK., Chan, JF., Lau, SK., Woo, PC. \& Yuen, KY. (2015). A systematic approach to novel virus discovery in emerging infectious disease outbreaks. The Journal of Molecular Diagnostics, 17(3), 230-241. DOI: 10.1016/j.jmoldx.2014.12.002

Stanley, MA. (2020). Epithelial cell responses to infection with human papillomavirus. Clinical Microbiology Reviews, 25(2), 215-222. DOI: 10.1128/cmr.05028-11

Stenlund, A. (2003). Initiation of DNA replication: Lessons from viral initiator proteins. Nature Reviews Molecular Cell Biology, 4, 777-785. DOI: 10.1038/nrm1226

Sterling, JC., Handfield-Jones, S. \& Hudson, PM. (2001). Guidelines for the management of cutaneous warts. British Journal of Dermatology, 144(1), 4-11. DOI: 10.1046/j.1365-2133.2001.04066.x

Stoler, MH., Wolinsky, SM., Whitbeck, A., Broker, TR. \& Chow, LT. (1989). Differentiation-linked human papillomavirus types 6 and 11 transcription in genital condylomata revealed by in situ hybridization with message-specific RNA probes. Virology, 172(1), 331-340. DOI: 10.1016/0042-6822(89)90135-9

Strauss, MJ., Shaw, EW., Bunting, H. \& Melnick, JL. (1949). "Crystalline" virus-like particles from skin papillomas characterized by intranuclear inclusion bodies. Sage Journals, 72(1), 46-50. DOI: 10.3181/00379727-72-17328

Swain, SK., Das, A., Sahoo, L., Debta, P. (2020). Pediatric laryngeal papillomatosis: A review. Annals of Indian Academy of Otorhinolaryngology Head and Neck Surgery, 4(2), 19-23. DOI: 10.4103/aiao.aiao_11_19

Taku, O., Businge, CB., Mdaka, ML., Phohlo, K., Basera, W., Garcia-Jardon, M., Meiring, TL., Gyllensten, U., Williamson, AL. \& Mbulawa, Z. (2020). Human papillomavirus prevalence and risk factors among HIV-negative and HIV-positive women residing in rural Eastern Cape, South Africa. International Journal of Infectious Diseases, 95, 176-182. DOI: 10.1016/j.ijid.2020.02.051

The International HPV Reference Centre. (2021). Human Reference clones, viewed on 22 July 2021, [https://www.hpvcenter.se/human_reference_clones](https://www.hpvcenter.se/human_reference_clones)

Tjon Pian Gi, RE., Dietz, A., Djukic, V., Eckel, HE., Friedrich, G., Golusinski, W., Hantzakos, A., Lawson, G., Remacle, M., Rihkanen, H. \& Dikkers, FG. (2012). Treatment of recurrent respiratory papillomatosis and adverse reactions following off-label use of cidofovir (Vistide ${ }^{\circledR}$ ). European archives of Otorhino-Laryngology, 269(2), 361-362. DOI: 10.1007/s00405-011-1804-7

Trifinopoulos, J., Nguyen, LT., von Haeseler, A. \& Minh, BQ. (2016) W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Research, 44, W232-W235. DOI: 10.1093/nar/gkw256

Tseng, CJ., Liang, CC., Soong, YK. \& Pao, CC. (1998). Perinatal transmission of human papillomavirus in infants: Relationship between infection rate and mode of delivery. Obstetrics \& Gynecology, 91(1), 92-96. DOI: 10.1016/S0029-7844(97)00594-2

Tseng, CJ., Lin, CY., Wang, RL., Chen, LJ., Chang, YL., Hsieh, TT. \& Pao, CC. (1992). Possible transplacental transmission of human papillomaviruses. American Journal of Obstetrics and Gynecology, 166(1 Pt 1), 35-40. DOI: 10.1016/0002-9378(92)91825-u

Tuna, M. \& Amos, CI. (2017). Next generation sequencing and its applications in HPV-associated cancers. Oncotarget, 8(5), 8877-8889. DOI: 10.18632/oncotarget. 12830
van Doorslaer, K. (2013). Evolution of the Papillomaviridae. Virology, 445,11-20. DOI: 10.1016/j.virol.2013.05.012
van Regenmortel, MHV., Fauquet, CM., Bishop, DHL., Calisher, CH., Carsten, EB., Estes, MK., Lemon, SM., Maniloff, J., Mayo, MA., McGeoch, DJ., Pringle, CR. \& Wickner, RB. (2002). Virus Taxonomy. Seventh Report of the International Committee for the Taxonomy of Viruses. Academic Press, New-York

Venuti, A., Paolini, F., Nasir, L., Corteggio, A., Roperto, S., Campo, MS. \& Borzacchiello, G. (2011). Papillomavirus E5: the smallest oncoprotein with many functions. Molecular Cancer, 10, 140. DOI: 10.1186/1476-4598-10-140

Wallace, NA. \& Galloway, DA. (2014). Manipulation of cellular DNA damage repair machinery facilitates propagation of human papillomaviruses. Seminars in Cancer Biology, 26, 30-42. DOI: 10.1016/j.semcancer.2013.12.003

Wang, JW. \& Roden, RB. (2013). L2, the minor capsid protein of papillomavirus. Virology, 445(1-2), 175-186. DOI: 10.1016/j.virol.2013.04.017

Watts, DH., Koutsky, LA., Holmes, KK., Goldman, D., Kuypers, J., Kiviat, NB. \& Galloway, DA. (1998). Low risk of perinatal transmission of human papillomavirus: Results from a prospective cohort study. American Journal of Obstetrics and Gynecology, 178(2), 365-373. DOI: 10.1016/s0002-9378(98)80027-6

Welschmeyer, A. \& Berke, G. (2021). An updated review of the epidemiological factors associated with recurrent respiratory papillomatosis. Laryngoscope Investigative Otolaryngology. 6. DOI: 10.1002/lio2.521.

Wilcox, LJ., Hull, BP., Baldassari, CM. \& Derkay, CS. (2014). Diagnosis and management of recurrent respiratory papillomatosis. The Pediatric Infectious Disease Journal, 33(12), 1283-1284. DOI: 10.1097/inf.0000000000000551

Wu, X., Zhang, C., Feng, S., Liu, C., Li, Y., Yang, Y., Gao, J., Li, H., Meng, S., Li, L., Zhang, Y., Hu, X., Wu, X., Lin, L., Li, X. \& Wang, Y. (2009). Detection of HPV types and neutralizing antibodies in Gansu province, China. Journal of Medical Virology, 81, 693-702. DOI: 10.1002/jmv. 21435

Yajid, AI., Zakariah, MA., Mat Zin, AA. \& Othman, NH. (2017). Potential role of E4 protein in human papillomavirus screening: A review. Asian Pacific Journal of Cancer Prevention, 18(2), 315-319. DOI: 10.22034/арјср.2017.18.2.315

Zahreddine, M., Mayrand, MH., Therrien, C., Trevisan, A., Dagenais, C., Monnier, P., Laporte, L., Niyibizi, J., Deshaies, C., Carceller, AM., Fraser, W., Brassard, P., Lacroix, J., Bédard, MJ., Girard, I., Audibert, F., Coutlée, F. \& Trottier, H. (2020). Antibodies to human papillomavirus types 6, 11, 16 and 18: Vertical transmission and clearance in children up to two years of age. EClinicalMedicine, 21, 100334. DOI: 10.1016/j.eclinm.2020.100334

Zhang, J., Zhang, S., Wang, M., Ding, X., Wen, Q., Chen, Z., Cao, M., Jing, Y. \& Zhang, S. (2018). Genetic variability in E5, E6, E7 and L1 genes of human papillomavirus type 31. Molecular Medicine Reports, 17(4), 5498-5507. DOI: 10.3892/mmr.2018.8500

Zouridis, A., Kalampokas, T., Panoulis, K., Salakos, N. \& Deligeoroglou, E. (2018). Intrauterine HPV transmission: A systematic review of the literature. Archives of Gynecology and Obstetrics, 298(1), 3544. DOI: 10.1007/s00404-018-4787-4
zur Hausen, H. (2002). Papillomaviruses and cancer: From basic studies to clinical application. Nature Reviews Cancer, 2, 342-350. DOI: 10.1038/nrc798
zur Hausen, H., Gissmann, L., Steiner, W., Dippold, W. \& Dreger, I. (1975). Human papilloma viruses and cancer. Bibliotheca Haematologica. 43, 569-71. DOI: 0.1159/000399220

## Appendices

Appendix 1: Human papillomavirus type 11 positive patient sample information

| Date of Birth | Sex | Diagnosis | VBD number |
| :---: | :---: | :---: | :---: |
| 2014/06/19 | F | 2018/03/14 | 114/18 |
| 2014/01/07 | F | 2018/02/27 | 12/18* |
| 2014/07/03 | F | 2018/02/15 | 113/18 |
| 2013/03/31 | M | 2017/07/26 | 23/17* |
| 1970/03/03 | M | 2017/06/21 | 47/17 |
| 1985/08/22 | M | 2017/06/13 | 21/17 |
| 1985/05/16 | M | 2016/12/05 | 68/16* |
| 2012/03/13 | M | 2016/10/28 | 48/16* |
| 2009/07/23 | M | 2016/08/17 | 33/16* |
| 2002/11/12 | M | 2016/01/24 | 17/16 |
| 2012/01/08 | F | 2015/11/18 | 68/15* |
| 1992/03/28 | F | 2015/10/23 | 70/15 |
| 2007/06/04 | F | 2015/09/16 | 49/15* |
| 2009/03/04 | M | 2015/09/16 | 50/15 |
| 1990/05/18 | M | 2015/09/08 | 41/15* |
| 2013/04/23 | F | 2015/09/01 | 44/15* |
| 2011/09/01 | M | 2015/08/19 | 43/15 |
| 2013/10/28 | M | 2015/05/06 | 16/15* |
| 1949/01/05 | M | 2015/04/15 | 08/15 |
| 2013/07/24 | F | 2015/04/03 | 14/15* |
| 2011/05/29 | M | 2014/08/02 | 41/14* |
| 2010/01/28 | M | 2014/07/25 | 38/14 |
| 1971/03/08 | F | 2014/07/23 | 36/14 |
| 2009/03/29 | M | 2014/06/26 | 28/14* |
| 2011/03/17 | M | 2014/05/27 | 17/14* |
| 2000/07/05 | M | 2014/05/26 | 22/14* |
| 2007/11/12 | M | 2014/03/18 | 29/14* |
| 2006/01/01 | F | 2013/12/09 | 55/13* |
| 2011/09/07 | F | 2013/10/19 | 37/13 |
| 2011/04/24 | M | 2013/08/02 | 29/13* |
|  |  | Continued... |  |


| Date of Birth | Sex | Diagnosis | VBD number |
| :---: | :---: | :---: | :---: |
| 2010/06/11 | F | 2013/01/25 | 08/13* |
| 2010/08/07 | M | 2012/10/07 | 48/12* |
| 2008/01/27 | M | 2012/09/26 | 47/12* |
| 2011/07/08 | M | 2012/07/31 | 39/12* |
| 2003/10/21 | F | 2012/06/12 | 37/12* |
|  |  |  | 14/14 |
| 2008/12/25 | M | 2012/01/03 | 09/12* |
| 1992/03/25 | F | 2011/11/30 | 61/11* |
|  |  |  | 62/11* |
| 1999/08/08 | M | 2011/11/02 | 58/11* |
| 2011/01/30 | M | 2011/09/19 | 52/11* |
| 1968/12/21 | M | 2011/08/15 | 41/11* |
| 2009/11/11 | M | 2011/07/25 | 37/11* |
| 2010/04/15 | F | 2011/07/04 | 35/11* |
| 2008/02/20 | M | 2011/06/14 | 49/12* |
| 2008/05/16 | M | 2011/05/23 | 30/11* |
| 2008/06/06 | M | 2011/03/07 | 34/11* |
| 2010/08/11 | F | 2011/02/26 | 14/11* |
|  |  |  | 18/14 |
| 2008/08/30 | F | 2010/12/29 | 04/11* |
| 2000/10/29 | F | 2010/12/10 | 20/16 |
| 2008/10/26 | M | 2010/10/13 | 48/10* |
| 2009/05/06 | M | 2010/08/18 | 33/10* |
|  |  |  | 63/11* |
|  |  |  | 01/12 |
|  |  |  | 02/12* |
|  |  |  | 03/12 |
|  |  |  | 04/12* |
| 1999/03/30 | F | 2010/07/26 | 26/10* |
| 2008/04/23 | F | 2010/06/30 | 21/10* |
| 2003/02/11 | M | 2010/06/11 | 23/10* |
| 2007/01/04 | M | 2010/01/06 | 01/10* |
|  |  |  | 07/10* |
| Continued... |  |  |  |


| Date of Birth | Sex | Diagnosis | VBD number |
| :---: | :---: | :---: | :---: |
|  |  |  | 17/10* |
|  |  |  | 34/10* |
| 1975/06/26 | M | 2009/11/23 | 79/09* |
| 2007/06/27 | M | 2009/10/23 | 74/09* |
|  |  |  | 15/10* |
| 1967/12/22 | M | 2009/06/09 | 59/09* |
|  |  |  | 63/09* |
|  |  |  | 69/15* |
| 2000/07/28 | M | 2008/08/05 | 55/08* |
| 2000/10/14 | M | 2008/01/14 | 34/08* |
| 1981/04/23 | F | 2006/12/13 | 01/08 |
| 2004/07/01 | F | 2005/11/28 | 04/08 |
|  |  |  | 37/08* |
|  |  |  | 08/09* |
|  |  |  | 14/09* |
| 2003/03/22 | F | 2005/10/27 | 41/12* |
| 2001/11/12 | F | 2003/10/01 | 17/09* |
| 1997/09/29 | M | 2001/10/09 | 49/08* |
|  |  |  | 59/08* |
|  |  |  | 69/09* |
|  |  |  | 16/10* |
| 1997/02/23 | M | 2000/06/19 | 16/09* |
|  |  |  | 52/09* |
|  |  |  | 37/10* |
| 1964/06/08 | F | 1997/06/30 | 33/11* |
| 1993/09/13 | M | 1997/04/29 | 05/08 |
|  |  |  | 45/08* |
|  |  |  | 13/10* |
| 1977/05/01 | F | 1990/07/15 | 69/16 |
| 1979/08/04 | F | 1988/08/05 | 54/08* |
| 1979/11/18 | F | 1984/11/18 | 77/09 |
| *Human papillomavirus type 11 samples included in this study |  |  |  |

Appendix 2: Human papillomavirus type 11 isolates retrieved from GenBank and accession numbers

| Isolate name on GenBank | GenBank <br> Accession number | Length of complete genome |
| :---: | :---: | :---: |
| M14119.1 Human papillomavirus type 11 (HPV-11) complete genome | M14119.1 | 7934bp |
| LN833187.1 Human papillomavirus type 11 complete genome, isolate 27 | LN833187.1 | 7934bp |
| LN833185.1 Human papillomavirus type 11 complete genome, isolate 25 | LN833185.1 | 7934bp |
| LN833184.1 Human papillomavirus type 11 complete genome, isolate 24 | LN833184.1 | 7934bp |
| LN833183.1 Human papillomavirus type 11 complete genome, isolate 23 | LN833183.1 | 7948bp |
| LN833169.1 Human papillomavirus type 11 complete genome, isolate 9 | LN833169.1 | 7934bp |
| LN833165.1 Human papillomavirus type 11 complete genome, isolate 5 | LN833165.1 | 7934bp |
| LN833161.1 Human papillomavirus type 11 complete genome, isolate 1 | LN833161.1 | 7934bp |
| KU298879.1 Human papillomavirus type 11 isolate 83A.11, complete genome | KU298879.1 | 7932bp |
| JQ773412.1 Human papillomavirus type 11 strain CU20, complete genome | JQ773412.1 | 7934bp |
| JQ773411.1 Human papillomavirus type 11 strain CU19, complete genome | JQ773411.1 | 7934bp |
| JQ773409.1 Human papillomavirus type 11 strain CU17, complete genome | JQ773409.1 | 7934bp |
| JQ773408.1 Human papillomavirus type 11 strain CU16, complete genome | JQ773408.1 | 7934bp |
| JN644141.1 Human papillomavirus type 11 isolate GUMC-AJ, complete genome | JN644141.1 | 7934bp |
| HE611263.1 Human papillomavirus type 11 complete genome, isolate LP220 | HE611263.1 | 7934bp |
| HE574702.1 Human papillomavirus type 11 complete genome, isolate JO-RRP_2 | HE574702.1 | 7934bp |


| FR872717.1 Human papillomavirus type 11 complete genome | FR872717.1 | 7934bp |
| :---: | :---: | :---: |
| Continued... |  |  |
| Isolate name on GenBank | GenBank <br> Accession number | Length of complete genome |
| FN907963.1 Human papillomavirus type 11 complete genome, isolate CS20 | FN907963.1 | 7943bp |
| FN907962.1 Human papillomavirus type 11 complete genome, isolate CAC86 | FN907962.1 | 7934bp |
| FN870021.1 Human papillomavirus type 11 complete genome, isolate A86 | FN870021.1 | 7934bp |
| EU918768.1 Human papillomavirus type 11 isolate LZod45-11, complete genome | EU918768.1 | 7934bp |
| MN788368.1 Human papillomavirus type 11 isolate JO-RRP10, complete genome | MN788368.1 | 7934bp |
| MK463921.1 Human papillomavirus type 11 isolate HPV11-gw1110, complete genome | MK463921.1 | 7932bp |
| MK463916.1 Human papillomavirus type 11 isolate HPV11-gw1105 , complete genome | MK463916.1 | 7932bp |
| MK463914.1 Human papillomavirus type 11 isolate HPV11-gw1103, complete genome | MK463914.1 | 7934bp |
| MK313767.1 Human papillomavirus type 11 isolate CAC1/HPV11, complete genome | MK313767.1 | 7934bp |
| MK313765.1 Human papillomavirus type 11 isolate JORRP8/HPV11, complete genome | MK313765.1 | 7934bp |
| MK313763.1 Human papillomavirus type 11 isolate JORRP2/HPV11, complete genome | MK313763.1 | 7934bp |



## Dear Ms Corne Thuynsma

Ethics Number: UFS-HSD2019/1109/2708
Ethics Clearance: Genetic analysis of human papillomavirus type 11 isolates from patients with recurrent respiratory papillomatosis treated at Universitas Academic Hospital
Principal Investigator: Ms Corne Thuynsma
Department: Medical Microbiology Department (Bloemfontein Campus)
SUBSEQUENT SUBMISSION APPROVED

With reference to your recent submission for ethical clearance from the Health Sciences Research Ethics Committee. I am pleased to inform you on behalf of the HSREC that you have been granted ethical clearance for your request as stipulated below:

- Continuation report and annual re-approval until 31/07/2021.

The HSREC functions in compliance with, but not limited to, the following documents and guidelines: The SA National Health Act. No. 61 of 2003; Ethics in Health Research: Principles, Structures and Processes (2015); SA GCP(2006); Declaration of Helsinki; The Belmont Report; The US Office of Human Research Protections 45 CFR 461 (for non-exempt research with human participants conducted or supported by the US Department of Health and Human Services- (HHS), 21 CFR 50, 21 CFR 56; CIOMS; ICH-GCP-E6 Sections 1-4; The International Conference on Harmonization and Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH Tripartite), Guidelines of the SA Medicines Control Council as well as Laws and Regulations with regard to the Control of Medicines, Constitution of the HSREC of the Faculty of Health Sciences.

For any questions or concerns, please feel free to contact HSREC Administration: 051-4017794/5 or email EthicsFHS@ufs.ac.za.
Thank you for submitting this request for ethical clearance and we wish you continued success with your research.

Yours Sincerely


Dr. SM Le Grange
Chair : Health Sciences Research Ethics Committee
Health Sciences Research Ethics Committee
Office of the Dean: Health Sciences
T: +27 (0)51 $4017795 / 7794$ | E: ethicsfhs@ufs.ac.za
IRB 00011992; REC 230408-011; IORG 0010096; FWA 00027947
Block D, Dean's Division, Room D104 | P.O. Box/Posbus 339 (Internal Post Box G40) |Bloemfontein 9300 | South Africa
www.ufs.ac.za

# Appendix 4: Design of E2 segment primers 

```
Forward primer 5'-ACAGTGCAGCTACGCCTATA-3' (Blue)
Reverse primer 5'-TTGTACAGGCACTACCTCCATAC-3' (Green)
>M14119.1 Human papillomavirus type 11 (HPV-11) complete genome
```

ACAGTGCAGCTACGCCTATAGTGCAACTGCAAGGTGATTCCAATTGTTTAAAATGTTTTAGATATAGACTGA ATGACAAATATAAACATTTGTTTGAATTAGCATCTTCAACGTGGCATTGGGCCTCACCTGAGGCACCACATAA AAATGCAATTGTAACATTAACATATAGCAGTGAGGAACAACGTCAGCAATTTTTAAACAGTGTAAAAATACCA CCCACCATTAGGCATAAGGTGGGGTTTATGTCATTACATTTATTGTAACCATTACACCTGTATATATGTATATG TGTACATAACATACGTGTATGGAGGTAGTGCCTGTACAA
>JN644141.1 Human papillomavirus type 11 isolate GUMC-AJ, complete genome

ACAGTGCAGCTACGCCTATAGTGCAACTGCAAGGTGATTCCAATTGTTTAAAATGTTTTAGATATCGACTGA ATGACAAATATAGACATTTGTTTGAATTAGCATCTTCAACGTGGCATTGGGCCTCACCTGAGGCACCACATAA AAATGCAATTGTAACATTAACATATAGCAGTGAGGAACAACGTCAGCAATTTTTAAACAGTGTAAAAATACCA CCCACCATTAGGCATAAGGTGGGGTTTATGTCATTACATTTATTGTAACCATTGCACCTGTATATATGTATATG TGTACATAACATACGTGTATGGAGGTAGTGCCTGTACAA
>JQ773408.1 Human papillomavirus type 11 strain CU16, complete genome
ACAGTGCAGCTACGCCTATAGTGCAACTGCAAGGTGATTCCAATTGTTTAAAATGTTTTAGATATCGACTGA ATGACAAATATAAACATTTGTTTGAATTAGCATCTTCAACGTGGCATTGGGCCTCACCTGAGGCACCACATAA AAATGCAATTGTAACATTAACCTATAGCAGTGAGGAACAACGTCAGCAATTTTTAAACAGTGTAAAAATACCA CCCACCATTAGGCATAAGGTGGGGTTTATGTCATTACATTTATTGTAACCATTGCACCTGTATATATGTATATG TGTACATAACATACGTGTATGGAGGTAGTGCCTGTACAA
>MN788368.1 Human papillomavirus type 11 isolate JO-RRP10, complete genome

ACAGTGCAGCTACGCCTATAGTGCAACTGCAAGGTGATTCCAATTGTTTAAAATGTTTTAGATATCGACTGA ATGACAAATATAGACATTTGTTTGAATTAGCATCTTCAACGTGGCATTGGGCCTCACCTGAGGCACCACATAA AAATGCAATTGTAACATTAACATATAGCAGTGAGGAACAACGTCAGCAATTTTTAAACAGTGTAAAAATACCA CCCACCTTAGGCATAAGGTGGGGTTTATGTCATTACATTTATTGTAACCATTGCACCTGTATATATGTATATGT GTACATAACATACGTGTATGGAGGTAGTGCCTGTACAA
>LN833187.1 Human papillomavirus type 11 complete genome, isolate 27
ACAGTGCAGCTACGCCTATAGTGCAACTGCAAGGTGATTCCAATTGTTTAAAATGTTTTAGATATAGACTGA ATGACAAATATAGACATTTGTTTGAATTAGCATCTTCAACGTGGCATTGGGCCTCACCTGAGGCACCACATAA AAATGCAATTGTAACATTAACATATAGCAGTGAKGAACAACGTCAGCAATTTTTAAACAGTGTAAAAATACC ACCCACCATTAGGCATAAGGTGGGGTTTATGTCATTACATTTATTGTAACCATTGTACCTGTATATATGTATAT GTGTACATAACATACGTGTATGGAGGTAGTGCCTGTACAA

## Appendix 5: Specificity of E2 segment primers

|  |  | Sequence ( $5^{\prime} \gg 3^{\text {a }}$ ) |  | Length | Tm | GC\% | Self complementarity | Self $3^{\prime}$ complementarity |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Forward primer |  | ACAGTGCAGCTACGCCTATA |  | 20 | 58.60 | 50.00 | 4.00 | 4.00 |
| Reverse primer |  | tTGtacaggcactacctccatac |  | 23 | 59.55 | 47.83 | 6.00 | 1.00 |
| Products on target templates |  |  |  |  |  |  |  |  |
| >MN788368.1 Human papillomavirus type 11 isolate JO-RRP10, complete genome |  |  |  |  |  |  |  |  |
| product length $=331$ |  |  |  |  |  |  |  |  |
| Forward primer | 1 | ACAGTGCAGCTACGCCTATA 20 |  |  |  |  |  |  |
| Template | 3561 | .................... 35 | 358 |  |  |  |  |  |
| Reverse primer | 1 | tTGtacaggcactacctccatac | 23 |  |  |  |  |  |
| Template | 3891 | ....................... | 3869 |  |  |  |  |  |
| >MK463922.1 Human papillomavirus type 11 isolate HPV11-gw-1111, complete genome |  |  |  |  |  |  |  |  |
| product length $=331$ |  |  |  |  |  |  |  |  |
| Forward primer | 1 | ACAGTGCAGCTACGCCTATA 20 |  |  |  |  |  |  |
| Template | 3460 | .................... 34 | 3479 |  |  |  |  |  |
| Reverse primer | 1 | tTGTACAGGCACTACCTCCATAC | 23 |  |  |  |  |  |
| Template | 3790 | ...................... | 3768 |  |  |  |  |  |
| >MK463921.1 Human papillomavirus type 11 isolate HPV11-gw-1110, complete genome |  |  |  |  |  |  |  |  |
| product length $=331$ |  |  |  |  |  |  |  |  |
| Forward primer | 1 | ACAGTGCAGCTACGCCTATA 20 |  |  |  |  |  |  |
| Template | 3460 | .................... 34 | 3479 |  |  |  |  |  |
| Reverse primer | 1 | TTGTACAGGCACTACCTCCATAC | 23 |  |  |  |  |  |
| Temolate | 3790 | ..................... | 3768 |  |  |  |  |  |
| >MK463920.1 Human papillomavirus type 11 isolate HPV11-gw-1109, complete genome |  |  |  |  |  |  |  |  |
| product length $=331$ |  |  |  |  |  |  |  |  |
| Forward primer |  | ACAGTGCAGCTACGCCTATA $2 \theta$ |  |  |  |  |  |  |
| Template | 3460 | .................... 34 | 3479 |  |  |  |  |  |
| Reverse primer Template | $\begin{aligned} & 1 \\ & 3790 \end{aligned}$ | tTGTACAGGCACTACCTCCATAC | $\begin{aligned} & \text { c } 23 \\ & \text { - } 3768 \end{aligned}$ |  |  |  |  |  |

```
l}\begin{array}{l}{\mathrm{ product length = 331 }}\\{\mathrm{ Forward primer 1 ACAGTGCAGCTACGCCTATA 20}}
Template 3460 3479
lllllal:
```

>MK463918.1 Human papillomavirus type 11 isolate HPV11-gw-1107, complete genome

| Forward primer | 1 | acagtgcagctacgectata | 28 |
| :---: | :---: | :---: | :---: |
| Template | 3460 |  | 3479 |
| Reverse primer | 1 | tTGTACAGGCACTACCTCCA | 23 |
| Template | 3798 |  | 3768 |

>MK463917.1 Human papillomavirus type 11 isolate HPV11-gw-1106, complete genome
product length $=331$
Forward primer 1 ACAGTGCAGCTACGCCTATA 2
Template 3460 ACAGGCAGCTACGCCTATA 20

| Seareh parameters and other details |  |
| :--- | :--- |
| Number of Blast hits analyzed | 3256 |
| Entrez query |  |
| Min total mismatches | 2 |
| Min 3' end mismatches | 2 |
| Defined 3' end region length | 5 |
| Mismatch threshold to ignore targets | 6 |
| Max target size | 4000 |
| Max number of Blast target sequences | 50000 |
| Blast E value | 30000 |
| Blast word size | 7 |
| Max candidate primer pairs | 500 |
| Min PCR product size | 63 |
| Max PCR product size | 1000 |
| Min Primer size | 15 |
| Opt Primer size | 20 |
| Max Primer size | 25 |
| Min Tm | 57 |
| Opt Tm | 60 |
| Max Tm | 63 |
| Max Tm difference | 3 |
| Repeat filter | AUTO |
| Low complexity filter | Yes |

## Appendix 6: 1x TAE preparation

## TAE Buffer 50x Stock Recipe

- 242 g tris base in double-distilled H2O
- 57.1 ml glacial acetic acid
- 100 ml 0.5 M EDTA solution ( pH 8.0 )

To make the 1 x TAE working buffer, add 49 parts of dH2O to 1 part of 50 x TAE buffer

Appendix 7: Agarose gel preparation

| 1\% Agarose gel |  |
| :---: | :---: |
| $\mathbf{4 0 m l ~ 1 \% T A E}$ | 0.4 g agarose powder |
| $\mathbf{5 0 m l ~ 1 \% T A E}$ | 0.5 g agarose powder |
| $\mathbf{1 0 0 \mathrm { ml } 1 \%} \mathbf{T A E}$ | 1.0 g agarose powder |
| $\mathbf{2 0 0 m l ~ 1 \% T A E}$ | 2.0 g agarose powder |

## Appendix 8: GelRed stain preparation

To make 50 ml GelRed stain:

- 10ul 10000 x GelRed® Nucleic Acid Gel Stain (Biotum) $+45 \mathrm{ml} \mathrm{dH} 20+5 \mathrm{ml} 0.1 \mathrm{M} \mathrm{NaCl}$
- Only use 2-3 times before preparing a new GelRed stain
- Store GelRed stain in the dark


## Appendix 9: Amino acid codon chart



Appendix 10: Pairwise analysis of percentage divergence of nucleotides using the human papillomavirus type 11 concatenated E5a/b-L1-URR data set


## Continued...



## Continued...



## Continued...



Appendix 11: Pairwise analysis of percentage divergence of nucleotides using the 208bp human papillomavirus type 11 E 2 segment data set



Continued...


Continued...


## Continued..



Appendix 12: Pairwise analysis of percentage divergence of nucleotides using the human papillomavirus type 11 complete genomes


Continued．．．

| 8L200＇0 | ${ }^{\text {to }}$（ $0^{\prime} 0$ | $91800{ }^{\circ} 0$ | ESZOO＇0 | 8LZ00＇0 | 8L200＇0 | 99200＇0 | 16200＇0 | 8LZ00＇0 | 9St00＇0 | $162000^{\circ}$ | $99200{ }^{\prime} 0$ | 8LZ00＇0 | 8LZOO＇0 | $992000^{\prime}$ | $99200{ }^{\circ}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 乙SLOO＇0 | ャ9100＇0 | 92100＇0 | SZ000＇0 | 92000＇0 | $88000{ }^{\circ}$ | $88000{ }^{\circ} 0$ | 9 $2000{ }^{\circ} 0$ | 乙8ヶ00＇0 | $88000{ }^{\circ} 0$ | $88000{ }^{\circ}$ | 乙SLOO＇0 | 2SLOO＇0 | E1000＇0 | ع1000＇0 |
|  |  | E9000＇0 | 10100＇0 | ZS100＇0 | ZS $100{ }^{\circ}$ | t9100＇0 | ャ9100＇0 | 己SLOO＇0 | LOS00＇0 | 79100＇0 | $68100{ }^{\text {c }}$ | 9Z100＇0 | 92100＇0 | 681000 | $68100{ }^{\circ}$ |
|  |  |  | カト100＇0 | t9100＇0 | 79100＇0 | LLLOOO 0 | LLLOO＇0 | 79100＇0 | OZS00＇0 | LLLOOO 0 | 乙S $100{ }^{\text {O }}$ | $68100{ }^{\text {a }}$ | 6E100＇0 | ZS100＇0 | ZSLOO＇0 |
|  |  |  |  | $92100{ }^{\circ}$ | $9 \mathrm{P} 100{ }^{\circ}$ | $681000^{\circ}$ | $681000^{0}$ | $92100{ }^{\circ}$ | 9St00＇0 | $68100{ }^{\circ}$ | จLLOO＇0 | $92000{ }^{\circ}$ | 9L000＇0 | th $100{ }^{\circ}$ | †HLOO＇0 |
|  |  |  |  |  | $00000{ }^{\circ}$ | $88000{ }^{\circ}$ | $88000{ }^{\circ}$ | 9 $2000{ }^{\circ} 0$ | 乙8ヶ00＇0 | $88000{ }^{\circ}$ | $88000{ }^{\circ}$ | 乙SL00＇0 | 2S100＇0 | EL000＇0 | ع1000＇0 |
|  |  |  |  |  |  | $88000{ }^{\circ}$ | $880000^{\circ}$ | ¢ $2000{ }^{\circ}$ | 28ヶ00＇0 | $88000{ }^{\circ}$ | $88000{ }^{\circ}$ | 乙SLOO＇0 | 己St00＇0 | E1000＇0 | E1000＇0 |
|  |  |  |  |  |  |  | ¢ $20000^{\circ}$ | $88000{ }^{\circ}$ | 七6ヶ00＇0 | L9000＇0 | LS000＇0 | ¢9100＇0 | ャ9100＇0 | Sz000＇0 | ¢ $2000{ }^{\circ}$ |
|  |  |  |  |  |  |  |  | $88000{ }^{\circ}$ | ヤ6ヶ00＇0 | LS000＇0 | 15000＇0 | ャ9100＇0 | ャ9100＇0 | Sz000＇0 | GZ000＇0 |
|  |  |  |  |  |  |  |  |  | 69t00＇0 | $88000{ }^{\circ}$ | $88000{ }^{\circ}$ | ZSLOO＇0 | 2SLOO＇0 | ع1000＇0 | E1000＇0 |
|  |  |  |  |  |  |  |  |  |  | ＋6700＇0 | 69＋00＇0 | 76ヶ00＇0 | ャ6ヶ00＇0 | 69t00＇0 | 69t00＇0 |
|  |  |  |  |  |  |  |  |  |  |  | LS000＇0 | t9100＇0 | ャ9100＇0 | Sz000＇0 | 92000＊0 |
|  |  |  |  |  |  |  |  |  |  |  |  | $68100{ }^{\circ}$ | 68100＇0 | sz000＇0 | 9 20000 |
|  |  |  |  |  |  |  |  |  |  |  |  |  | 00000＇0 | $68100{ }^{\circ}$ | 68100＇0 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | 681000 | 6と100＇0 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 00000＇0 |

