Systematic Kaposi-Sarcoma Herpesvirus (KSHV) Genome Analysis from a South African HIV- and KSHV-positive patient cohort.

Melissa Thomas

Supervisor:

Dr Georgia Schäfer

Thesis presented for the degree of

Master of Science in Medicine

Medical Biochemistry

In the Division of

Medical Biochemistry and Structural Biology

Department of Integrative Biomedical Sciences

Faculty of Health Sciences

University of Cape Town

March 2024

DECLARATION

I, …MELISSA THOMAS……, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: ……11/03/2024…….

Acknowledgements

This project came together through the collaborative effort of a number of magnificent people. Firstly, I would like to thank my supervisor, Georgia Schäfer, for giving me this opportunity to not only assist in the KSHV research efforts, but for the countless learning opportunities and experiences I was given along the way. Thank you for lending a listening ear to my concerns and queries, and for your consistent encouragement and resolve over the past two years. The immeasurable support you have provided as a supervisor and the skills I have learnt will go far in my journey as a scientist.

I would like to thank our collaborators from the Whitby lab (NCI at Frederick, Frederick, MD, USA). Namely Vickie Marshall for her continued guidance and support over the past two years. Her patience and kindness in training both in person and online has meant the world to me. I would also like to thank Charles Goodman for his continued bioinformatic and computational support, and for putting time aside to help me with all my questions and issues whenever they arose. Charlie went above and beyond, and I am so thankful I was able to learn from him. It was a pleasure working with both of them and I would jump at the opportunity to do so again. I would like to also thank Elena Cornejo-Castro for answering my countless questions and lending me her knowledge. Thank you for reviewing my sequences time and time again as well as giving me direction when I needed it. I also want to thank both her and her husband for opening their home up to me during my visit to the lab towards the end of 2022. Their hospitality during my visit made the experience such a special one. I would also like to thank Joseph Meyer (Scientific Publications, Graphics, & Media, NCI at Frederick, Frederick, MD, USA) for his assistance with the phylogenetic tree graphics. Importantly, I would like to thank Denise Whitby for joining the collaborative effort in sequencing samples from South Africa and for working alongside the Schäfer lab for the past few years to grow KSHV research. I would like to thank her for allowing me the opportunity to train and carry out my research within her lab at Frederick, Maryland. The experience was immeasurable. To those also in the Whitby lab, Nazzarena Labo, Kyle Moore, Wendell Miley, and Romin Roshin, thank you for making my time there so special. I hold the memories dearly.

In the Schäfer Lab, I would like to thank Melissa Blumenthal and Humaira Lambarey for lending me their research samples to carry out this project and assisting me when necessary. I would

like to thank the Schäfer lab as a whole for being a group of remarkable individuals who support each other with kindness and humility. Thank you for being there for me during this research experience.

Lastly, I would like to thank The European and Developing Countries Clinical Trials Partnership (EDCTP) for funding my research and providing me with financial stability during the past two years. I would like to thank the Medical Research Council (MRC) and Poliomyelitis Research Foundation (PRF) for funds received to help with my student expenses. I would also like to thank the National Research Foundation (NRF) for the travel funding I received to help carry out my training in the United States. This project would not have been possible without this combined financial support.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIDS – Acquired Immunodeficiency Syndrome

ART – Anti-retroviral Therapy

Bp – base pairs

CT – Computed Tomography

DC-SIGN – Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin

EBV – Epstein Bar Virus

ERK 1 / 2 – Extracellular Signal-regulated Kinase

GUG – Gugulethu

HCMV – Human Cytomegalovirus

HHV – Human Herpesvirus

HIV – Human Immunodeficiency Virus

HREC – Human Research Ethics Committee

HSV – Herpes Simplex Virus

IL – interleukin

KDHTB – Khayelitsha Day Hospital Tuberculosis

KICS – KSHV-associated inflammatory cytokine syndrome

KS – Kaposi Sarcoma

KSHV – Kaposi Sarcoma Herpesvirus

LANA – Latency-associated Nuclear Antigen

MRI – Magnetic Resonance Imaging

MCD – Multicentric Castleman's disease

MCP – Major Capsid Protein

miRNA – micro-Ribonucleic Acid

MSM – Men who have Sex with Men

NFW – Nuclease-free Water

NFκβ – NF Kappa Beta

NGS – Next Generation Sequencing

ORF – Open Reading Frame

PCR – Polymerase Chain Reaction

PEL – Primary Effusion Lymphoma

RTA – Replication and Transcription Activator

SNP – Single Nucleotide Polymorphism

SSA – Sub-Saharan Africa

TB – Tuberculosis

TNFα – Tumour Necrosis Factor alpha

UCT – University of Cape Town

vFLIP – Fas-associated protein with death domain (FADD)-like interleukin-1β-converting enzyme (FLICE)-like inhibitory protein

vGPCR – viral G-Protein Coupled Receptor

vIRF – viral Interferon Regulatory Factor

VL – Viral Load

VR – Variable Region

VZV – Varicella-Zoster Virus

xCT – Cystine-glutamate Transporter

Abstract

Kaposi sarcoma herpesvirus (KSHV) is an oncogenic virus and the etiological agent of Kaposi's sarcoma, the most common AIDS-associated cancer. KSHV infection occurs primarily in Sub-Saharan Africa (SSA) where it is also associated with other malignancies such as Multicentric Castleman's Disease, Primary Effusion Lymphoma and KSHV-associated inflammatory cytokine syndrome, which also occur primarily with HIV co-infection. IN SSA most individuals acquire KSHV infection during childhood where it remains undetected until later in life when the burden of HIV takes its effect on the immune system.

There are 6 predominant KSHV subtypes world-wide, namely A, B, C, D, E and F which are determined by the highly variable K1 region of the KSHV genome. Sequencing and characterizing the circulating subtypes in a specific geographical region assist in tracking evolutionary changes as well as malignant outcomes associated with different subtypes. The aim of this study was to determine the circulating KSHV subtypes in the Western Cape region of South Africa. A total of 57 DNA samples isolated from peripheral blood of confirmed HIV/KSHV co-infected patients were selected according to KSHV viral load (VL) (> 5 copies/10 µL) and volume (> 5 µL). Of these, 20 were successfully Sanger sequenced to determine K1 subtype. Additionally, 9 samples met the criteria for whole-genome sequencing using Next Generation Sequencing (NGS). This study produced 29 K1 sequences (20 via Sanger and 9 via NGS) of which 26 were of the A subtype, specifically A5, and the remaining 3 were of the B subtype, namely B2. These results are consistent with previous studies from the same region where a trend of high A5 and B subtypes in AIDS-KS patients were reported. ORF-K15 is another variable and lytic gene associated with KSHV subtyping (alleles M, N and P). Sequencing results of the K15 gene, revealed two P subtypes, three M subtypes and four N subtypes in our samples. These novel results from the Western Cape of South Africa point towards an interesting distribution of subtypes contributing to previous reports of K15 P and M subtypes being the most prevalent in Africa.

There are limited whole-genome sequences available for South Africa, therefore these sequences provide a significant contribution to the pool of sequences available for this region.

1. Introduction

1.1 Kaposi Sarcoma Herpesvirus and Kaposi's Sarcoma

Kaposi Sarcoma Herpesvirus (KSHV), or human herpesvirus 8 (HHV-8), is the causative agent of Kaposi Sarcoma (KS), a vascular cancer originating from infected endothelial cells [13]. KS lesions primarily occur on the skin, however infection has been found to spread to mucosal membranes and visceral organs, such as the liver, lung, spleen and gastrointestinal tract [14]. KS was first described as a blood vessel tumour by the Hungarian dermatologist Moritz Kaposi in 1872, but in 1994, Chang *et al* discovered the etiological agent, KSHV, within KS lesions of individuals presenting with HIV-AIDS [13]. Although approximately 25% of all human cancers have been found to be etiologically tied to an infectious agent [14] specifically in low- to middle – income countries, KS is most often observed in those with decreased immunity and thus have a greater risk of developing the cancer. KS is seen in varying groups globally, and as such, has been divided into 4 main classes: Epidemic/AIDS-related KS, Classic/Mediterranean KS, Endemic/African KS and Iatrogenic/Transplant-related KS. Today, KS is largely understood as an AIDS defining malignancy and is one of the most prevalent cancers in Sub-Saharan Africa [15]. Transmission of this virus varies per geographic region according to behaviours, with men who have sex with men (MSM) transmission being higher in the US, notably after the AIDS outbreak in the 1980s. Nonsexual transmission is greater in African countries [16-18], where lifelong infection is typically established through transmission of saliva, such as from mother to child [19]. KS lesions can be diagnosed through a skin biopsy followed by anatomopathological examination. Lesions in difficult to see locations can be found via imaging (such as a Computed Tomography (CT) scan or Magnetic Resonance Imaging (MRI)), bronchoscopy or endoscopy. After cancer stage diagnosis, treatment centres around improving immune system function with the addition of chemo- and/or immunotherapy [20, 21]. KS survival rates vary based on numerous factors, namely cancer stage, the individual's age, general health and treatment plan. The current 5-year relative survival statistics for KS worldwide are: 81% for localized cancer, 65% for regional cancer and 47% for metastasized cancer [22]. In 2020, 34 270 new cases of KS were reported worldwide, with 15 086 people having died from KS globally in the same year [23].

KSHV infection also causes B cell lymphoproliferative disorders such Primary Effusion Lymphoma (PEL) [24], Multicentric Castleman disease (MCD) [25], and KSHV Inflammatory Cytokine Syndrome (KICS) [26] which is commonly observed alongside HIV infection (see 1.4 KSHV-associated malignancies). Anti-retroviral therapy (ART) has greatly decreased prevalence of epidemic KS; however, it is still the leading cancer in the context of AIDS [27, 28].

1.2 KSHV Structure and Genome

KSHV is the 8th member of the *Herpesviridae* family, joining Epstein-Barr virus (EBV), herpes simplex 1 (HSV-1) and 2 (HSV-2), HHV6, HHV7 as well as human cytomegalovirus (HCMV) and varicella-zoster virus (VZV). KSHV belongs to the sub family of *Gammaherpevirinae,* or gammaherpesviruses, alongside oncogenic EBV. To date it is the only human *Rhadinovirus* in existence [29]. Several studies have found relations to other Rhadinoviruses such as H*erpesvirus saimiri* in squirrel monkeys, H*erpesvirus ateles* in spider monkeys and other KSHV-related Rhadinoviruses infection of macaques, African green monkeys and chimpanzees [30-36]. The viral structure is typical for members of the *Herpesviridae* family and is made up of four layers consisting of a large core double-stranded DNA genome, enclosed by an icosapentahedral capsid surrounded by a tegument protein and a lipid envelope [37] (Figure 1). The viral particle consists of four structural proteins, namely MCP (major capsid protein), TR-1/2 (triplex component 1 and 2), SCIP and CSAF are integral for virion assembly. The tegument contains both inner and outer layers and consists of another 8 proteins (ORF 11, 21, 33, 45, 52, 63, 64 and 75) which may contribute towards genome entry and viral replication during early infection [38-41]. Situated within the viral envelope are eight glycoproteins, namely gB, K8.1 (A and B), gH, gL, gM, gN, ORF28 and ORF68, which are involved in host cell interaction [12, 39, 40, 42-45].

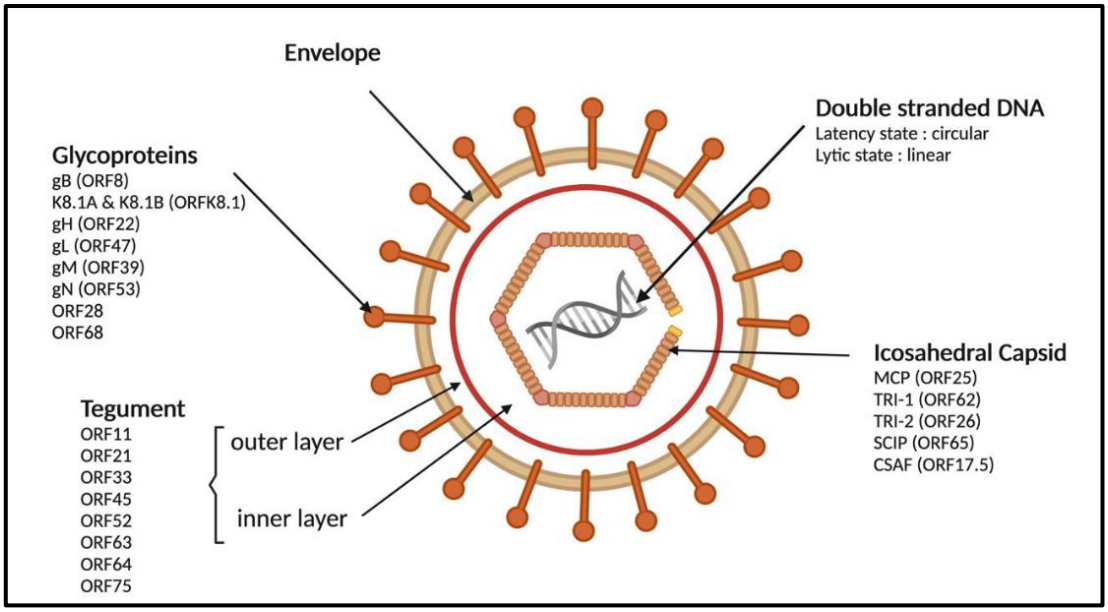


Figure 1: Structure of the Kaposi's sarcoma herpesvirus virion particle. The KSHV particle is composed of four distinct components, namely, the double stranded viral genome enclosed in a protein rich icosahedral capsid, electron dense tegument and lipid envelope. DNA: Deoxyribonucleic Acid; gX: Glycoprotein; MCP: Major Capsid Protein; ORF: Open Reading Frame; SCIP: Small Capsomer-Interacting Protein; TRI 1 and 2: Triplex Component 1 and 2. Created with BioRender.com [12].

KSHV virions contain a linear and largely conserved double stranded DNA genome approximately 165 - 179 kb in length. Of this, around 85% encodes 80 - 90 viral proteins (ORF4 -75 and K1 - K15) and 25 non-coding miRNAs (Figure 2). The highly variable 5' and 3' ends make up the remainder, which also serve as a basis for subtyping different strains. The first established KSHV genome, BC-1, was derived from a PEL cell line and exhibited a 165 kb dsDNA genome [29]. The first fully sequenced and annotated genome published in 2006, NC_009333.1 (GK18), displayed 86 genes, of which 22 were immunomodulatory [29, 46]. This sequence still serves as a reference sequence today, providing a basis for alignment and annotation of new whole-genome KSHV sequences.

As mentioned, higher gene variability occurs on the 5' and 3' end of the viral genome where the K1 and K15 genes are situated (Figure 2). As these genes are vital in oncogenesis and highly variable, they have been chosen for the characterization of viral KSHV subtypes. ORF K1 is found within the first open reading frame of the 5' end of the viral genome and codes for a transmembrane glycoprotein critical for tumorigenesis during the lytic life cycle. The K15 gene, situated at the 3' end, encodes an integral transmembrane protein vital for endothelial cell proliferation and migration during the lytic cycle [47]. K15 is involved in angiogenesis through early viral gene expression and virion production via activation of signalling pathways PLCγ1 and ERK1/2. It has been suggested that K15 could provide a target in therapeutic intervention due to its role in KSHV lytic replication and pathogenesis [48]. It is important to note

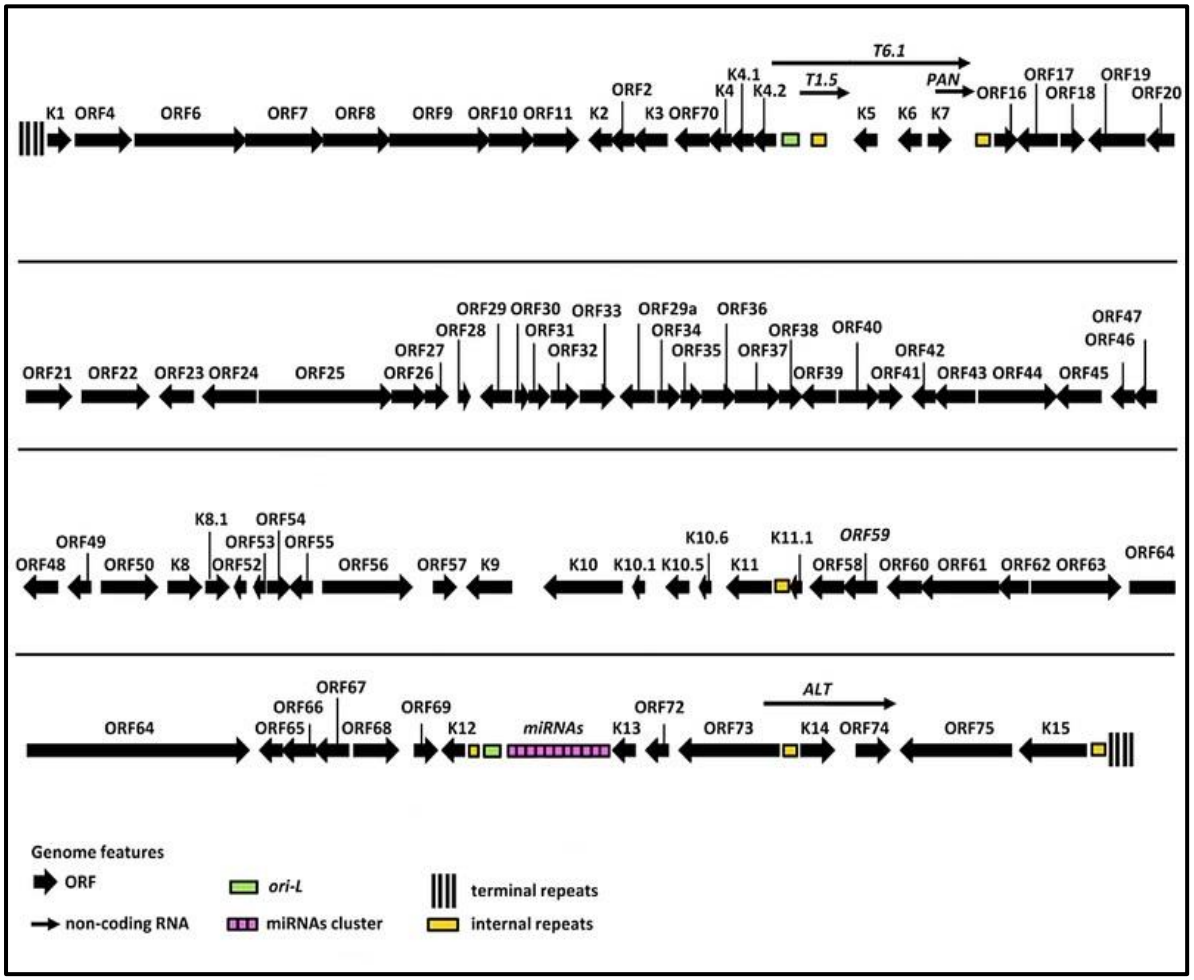


Figure 2: Linear genome map of KSHV. The genome begins with the K1 gene at the 5' end and ends with K15 at the 3' end. Arrows indicate gene sequence direction. ORF: Open Reading Frame; RNA: Ribonucleic acid; Ori-L: Origin of Lytic Replication; miRNA: microRNA. Adapted from Majerciak *et al* [7].

that the KSHV genome contains a number of genes, known as K1 through to K15, that are unique to this virus [49].

Within the central region of the KSHV genome lies a large portion of oncogenic proteins such as LANA-1 (Latency-Associated Nuclear Antigen 1), vFLIP (Fas-associated protein withdeath domain (FADD)-like interleukin-1β-converting enzyme (FLICE)-like inhibitory protein), v-cyclin, vGPCR (viral G-Protein Coupled Receptor), vIL-6 (viral Interleukin 6), vCCL, vMIP, vIRF (viral Interferon Regulatory Factor), etc. These proteins modify cellular pathways inducing common characteristics of cancer such as inhibition of apoptosis, increased cellular proliferation, inflammation and evasion of the immune system [12]. The central region of the KSHV genome is highly conserved with far less variation compared to the K1 and K15 genes [50-52]. KSHV is an ancient human virus that has co-evolved with its human host over millions of years as is typical for herpesviruses [50].

1.3 KSHV Infectious Cycle

Like other herpesviruses, KSHV, establishes latency within infected cells in order to evade the immune system and replicate successfully once the lytic cycle is activated (Figure 3). The aim is to produce viral progeny and prevent damage that would result in a compromise of viral fitness.

The entry of KSHV begins via interaction of its surface glycoproteins with the target cell's specific surface receptors (such as heparan sulphate, certain integrins, the xCT or Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN), EphA2 and other Eph receptors, depending on the cell type [53]), inducing endocytosis pathways and leading to uptake of the concentrated viral particles [54-57]. The viral DNA subsequently attaches to the host chromosome as an episome via tethering to LANA and hijacks the cells epigenetic machinery. From here, a heterochromatin structure is produced and is responsible for restricting viral gene expression with the aid of low levels of replication and transcription activator (RTA), an early lytic protein, and the master organizer, LANA [58]. Therefore, early lytic gene activity is used in establishing latency, but latency can however be achieved through transactivation of the LANA promoter by host transcription factors alone [1, 59]. Latency is

maintained through host signalling pathways, chromatin structure and selective gene expression. LANA plays the chief role in promoting repressive histone modifications and regulatory hoops in the DNA, with its ally vFLIP, also aiding in the repression of lytic activity [1]. Although lytic gene expression is silenced for the most part, minimal expression is maintained and lies in wait for reactivation through changes in RTA- and LANA- related activities [1]. The virus persists by replication of the host cell and segregates using the host cell machinery, transforming from extrachromosomal viral episomes into daughter "minichromosomes or episomes", where it remains in a quiescent state until lytic reactivation and viral replication [1, 60]. Once latency is established, KSHV infection becomes incurable and is managed through treatment aimed at inhibiting viral replication during the lytic cycle [1].

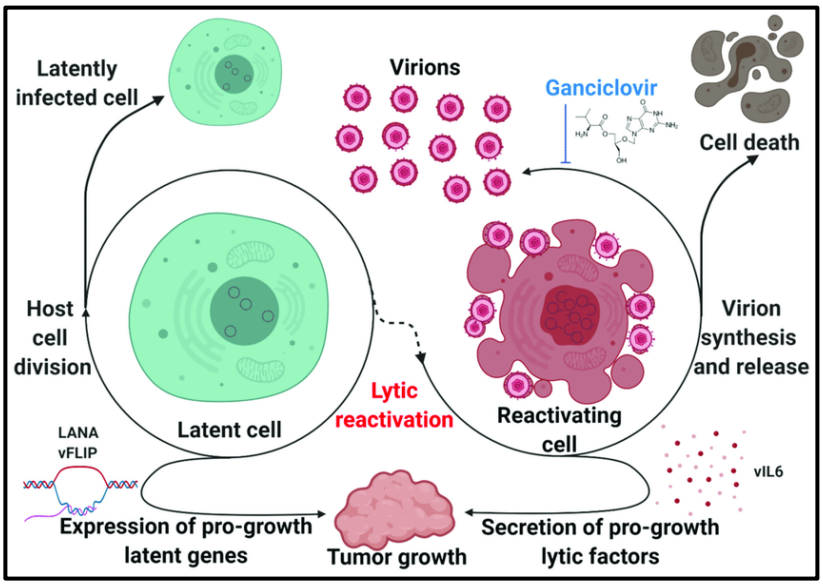


Figure 3: Latent and lytic cells contribution to tumour growth. The above figure illustrates the roles of both latent and lytic KSHV gene products in contributing to malignant transformation. Ganciclovir is an anti-viral drug. Figure adapted from Broussard and Damania [1].

KSHV reactivation can be achieved through RTA which leads to downstream transcription of lytic genes and later, viable virions [61]. Various physiological factors can result in activation of RTA, leading to promoter binding and expression, which subsequently transactivates other viral proteins [62-64]. Once the lytic cycle is activated, virion synthesis begins and mature virions are released for infection of nearby cells, leaving behind previously infected cells for subsequent cell death. Secretion of pro-growth factors stimulate tumorigenesis, as well as refresh latent supplies (Figure 3).

As previously mentioned, a unique balance between latent and lytic cycles is most desirable for KSHV's survival. Success is based on the virus successfully entering host cells, establishing life-long latency and reactivating to propagate virus production and transmission. Lytic activation and virion replication is crucial to transitioning from long-term latency to disease sites and viral spread to new hosts. Various studies have looked into what physiological and environmental impacts are responsible for the switch from latency to lytic cycles. These include oxidative stress, hypoxia, immunosuppression, imbalanced inflammatory cytokines and even viral co-infections. Not only have HIV and human papillomavirus shown to be responsible for KSHV reactivation, but increase of inflammatory cytokines from other members of the herpesvirus family, such as Herpes Simplex Virus (HSV) and HHV6, play a role as well [65, 66]. In contrast, co-infection with EBV has been reported to inhibit KSHV lytic replication [67].

It has been shown that both K1 and K15 play an integral role in lytic replication within the viral genome [68]. On stimulation, K1 activates a plethora of signalling molecules and activate numerous pathways during lytic activation and has been found to induce cell migration and angiogenesis [69-73]. K15, an important transmembrane protein, has also been found to be involved in endothelial cell angiogenesis and activation of pathways critical for lytic reactivation [48, 74].

1.4 KSHV-associated Pathologies

KSHV-associated malignancies are most commonly observed in immune compromised individuals; however, cancer development has been observed in the absence of overt immune deficiency. Malignancies can develop as a result of viral oncogene expression, and include cancers such as KS, PEL, MCD and KICS, which may present singularly or simultaneously. The KSHV lytic and latent life cycles play critical roles in attributing towards pathogenesis of these diseases as described in Figure 4.

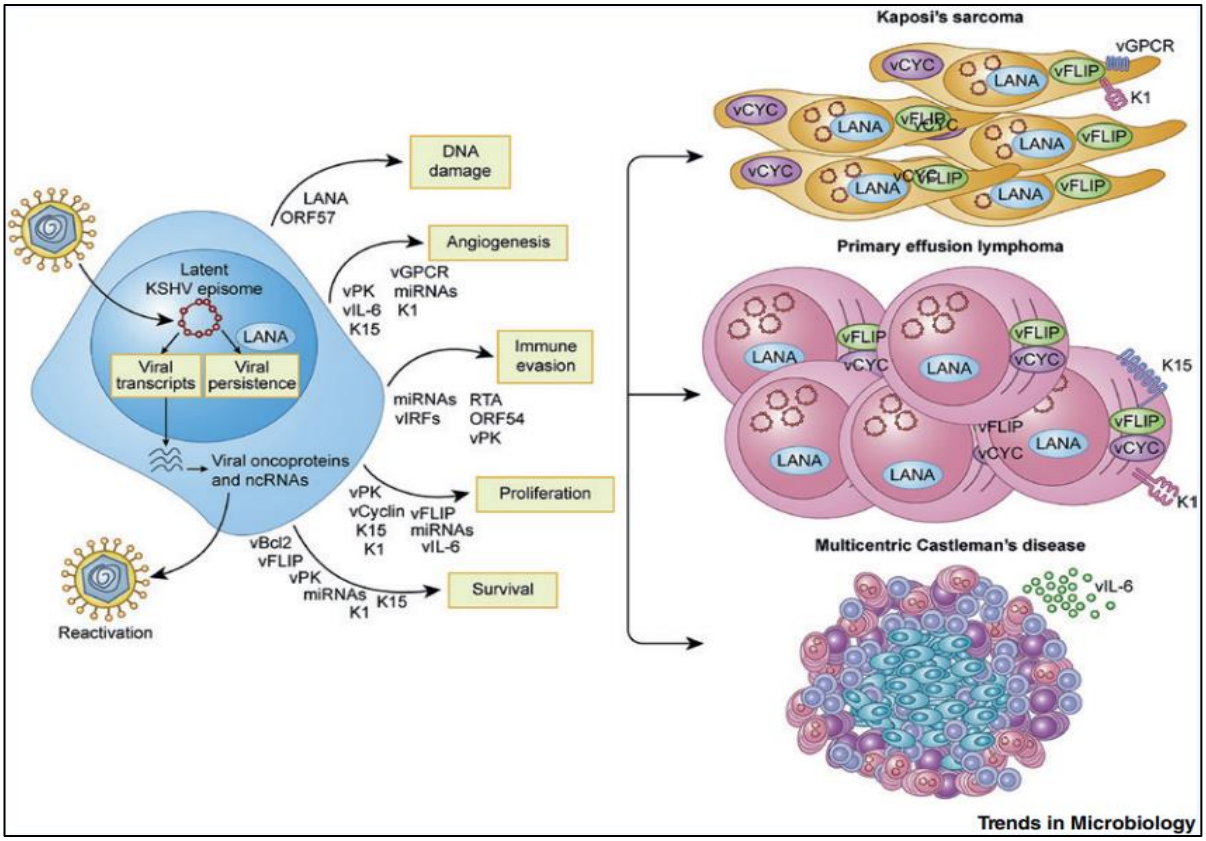


Figure 4: Establishment of KSHV within infected cells and potential malignancy outcomes. This illustration portrays viral entry and subsequent viral persistence with the aid of latent protein, LANA. On activation, viral transcripts are produced, coding for viral oncoproteins and non-coding RNAs (ncRNAs) establishing a state of lytic reactivation. Numerous signalling molecules, such as vIL-6, vIRFs and Rta, aid in DNA damage, angiogenesis, immune evasion, cell proliferation and survival, all leading to downstream tumour growth in the forms of KS, PEL and MCD. Note: KICS is not included in this figure. LANA: Latency-Associated Nuclear Antigen; ORF: Open Reading Frame; DNA: Deoxyribonucleic Acid; vIL-6: viral Interleukin-6; vIRFs: viral Interferon Regulatory Factors. Picture taken from Lange and Damania [5].

1.4.1 Kaposi Sarcoma

KSHV was first discovered in KS lesions from an AIDS infected individual, and became since known as the etiological agent for KS [13]. In later stages of the disease, the majority of cells are spindle shaped, with 90% containing the KSHV virus in its latent stage. Together with the small proportion responsible for lytic replication, these cells play important roles in KS pathogenesis [75-77]. KS lesions present as dark bluish-red or dark purple in colour and can progress to plaques and nodules [78, 79], as depicted in Figure 5. They can occur singularly, in clusters or widespread. Lesions appear as a result of excess proliferation of endothelial cells

and/or fibroblasts [80, 81], and may appear on the lower limbs, oesophageal tract, skin and lymph nodes. Initially, lesions are painless; however, swelling or itching later occurs as a result of inflammation, lymphoedema or secondary infection. With increased growth, these lesions often lead to a disrupt of vital visceral organs and become life threatening.

Misdiagnosis is frequent, with the macroscopic appearance of lesions being confused for drug-erupted dermatitis, interstitial granuloma annulare, and diabetic ulcers amongst others [82-84]. Moreover, nodules occurring in the oropharyngeal tract, including the oesophagus, throat and mouth, may bleed and be misdiagnosed for tuberculosis (TB) in a clinical setting [85]. In South Africa, limited resources and failing infrastructure often results in late-stage diagnosis and poor prognosis [86]. The gold standard approach to diagnosis is biopsy and histopathological diagnosis, followed by a cytotoxic chemotherapy driven treatment plan and ART (in the case of HIV-positive individuals) [87].

Figure 5: Kaposi Sarcoma lesions on the lower leg. Case report of KS lesions situated on the lower leg of an HIV positive male from India. Image taken from Mehta *et al* [9].

Kaposi sarcoma can occur as a result of varying environmental triggers, such as co-infections with other highly prevalent pathogens (HIV and malaria) [88, 89] and organ transplant-associated immunosuppression. In the context of Africa, endemic KS, or African KS, has been reported to be linked to genetic risk factors such as sex and ethnicity [90]. Due to the increased prevalence of HIV on the continent, AIDS KS accounts for the majority of the KS cases reported in Africa today. Global incidence of KS reported in 2020 showed 73% of KS cases were from Africa, 8.8% in Europe, 8.3% in Latin America and the Caribbean, 6.4% in Asia and 3.3% in North America [23]. Of all global mortalities reported, an 86.6% of KS-related deaths came from the African continent [23]. The influx of antiretrovirals observed since 2001 has led to a

decline in AIDS-related morbidity and mortality, including KS incidence, yet both are still observed at alarmingly elevated levels. In 2021, southernmost countries such as Namibia, Botswana, Zimbabwe and South Africa exhibited an elevated HIV incidence of between 11.8 - 18.3 % (15 to 49 years) within their respective populations [91]. A South African population-based survey conducted in 2017, found that 14% of South Africans (all ages) were HIV positive. Of these, 62.3% were exposed to ART, with the majority being women [92]. HIV prevalence in South Africa has doubled in the last 20 years, suggesting that HIV and therefore KS is expected to rise [93]. This high HIV prevalence, alongside the high KS incidence and mortality rates reported, showcase the dire need for improved diagnosis and treatment of KS on the African continent.

1.4.2 Multicentric Castleman's Disease

KSHV-associated Multicentric Castleman's Disease (MCD) is a plasmablastic, B-cell lymphoproliferative disorder most often documented in individuals living with HIV [94]. Substantial inflammation in KSHV-MCD is observed as a result of the production of excess human cytokines mainly IL-6 and IL-10, as well as TNFα and IL-1 [95-97]. The latent protein, LANA, is also upregulated in infected cells [98] (Figure 6). Although Castleman's Disease can occur singularly and present as a unicentric lymph node mass, KSHV infection promotes lymph-node based inflammation at multiple loci in the human body via increased expression of viral IL-6 [99]. This can result in the presence of neoplasms consisting of spindle-shaped cells within tissues (Figure 6). Clinically, KSHV-MCD is characterized by intermittent, dysregulated inflammatory symptoms, oedema, lymphadenopathy and splenomegaly, and can be confirmed pathologically through lymph node biopsy [100]. KSHV-MCD differs from other KSHV-associated malignancies in that it occurs in λ-restricted plasmablasts which commonly occur in the mantle or marginal zone of involved lymph nodes [101]. KSHV-MCD can be difficult to diagnose due to its significant overlap with other infections and malignancies of the lymphatic

system, such as KS and PEL, and requires KSHV-specific staining of lymph node biopsies. Symptoms include fatigue, fevers, night sweats, rashes, gastrointestinal and respiratory issues.

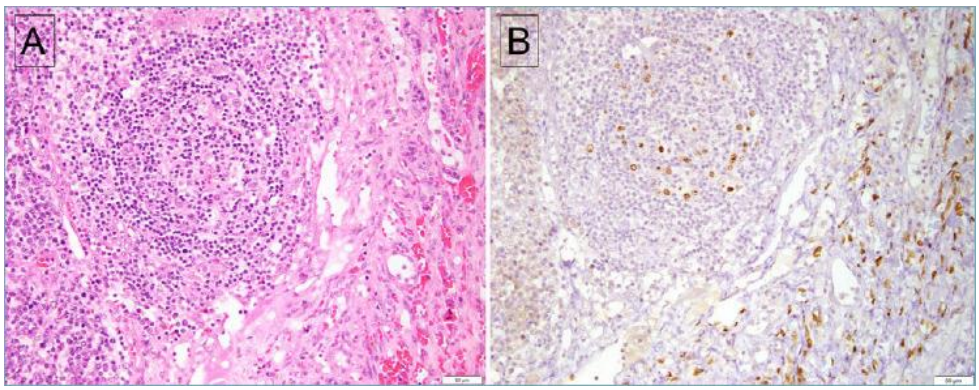


Figure 6: Stain images of a lymph node exhibiting both MCD and KS. A) H&E stain illustrating spindle cell proliferation in association with extravasated erythrocytes. B) LANA-positive stain in both lymphoid and spindle cells. Image taken from Zhou *et al* [11].

KSHV-MCD can advance to multi-organ failure and is fatal if left untreated with an average survival rate of 2 years. If caught early and given correct treatment, KSHV-MCD prognosis is promising with an overall survival rate of 71% for 10 years post-diagnosis having been reported [101].

1.4.3 Primary Effusion Lymphoma

Primary effusion lymphoma (PEL) is a distinctive form of non-Hodgkin lymphoma associated with KSHV. This lymphoproliferative disorder is characterised as an aggressive lymphoma of malignant pleural, pericardial and peritoneal effusions with the absence of a neoplasm (Figure 7). Infected cells are of B-cell origin whereby viral IL-6 is significantly more expressed than other lytic genes [102, 103]. PEL was first observed in the late 1980s during the AIDS epidemic, and was found to be associated with AIDS-KS [13, 104]. Thereafter, it was discovered that KSHV was the etiological agent of the lymphoma. Although EBV infection occurs frequently in PEL cases, it is not required for pathogenesis and may play an assistive role in the disease [105]. KSHV has been determined to be in the latent state in most PEL tumour cells and therefore evades any antiviral agents responsible for inhibiting lytic replication of the virus. LANA ensures latent replication of the viral episome, whilst simultaneously disrupting tumour suppressor proteins p53

and Rb [106, 107]. Viral cyclin is over-expressed in latent cells to aid in unregulated cellular proliferation [108-110], during which time vFLIP inhibits cell death by activating transcription of NF-κβ and preventing caspase 8 activation [111-113] (Figure 4).

PEL symptoms include fever, weight loss and abdominal discomfort due to internal extrinsic compression. There is no set treatment for PEL and prognosis remains poor even after chemotherapy [114].

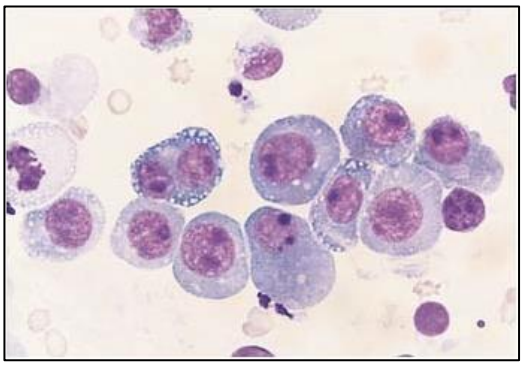


Figure 7: Wright-Giemsa stain illustrating pleural fluid containing large plasmacytoid tumour cells in a patient with PEL. Cytospin preparation at x1000 magnification. Image taken from Jones *et al* [8].

1.3.5 KSHV-associated Inflammatory Cytokine Syndrome

KSHV inflammatory cytokine syndrome (KICS) is another complication observed by KSHV infection and is associated with elevated levels of IL-6 and IL-10. It is described as 'MCD-like systemic inflammation in the absence of pathologic evidence of MCD' [115]. This disease displays a higher viral load in both HIV and non-HIV infected individuals [116] when compared to MCD. Histopathological features include hyperplasia in lymph nodes as well as shrinkage in KSHV infected plasma cells as a result of plasmolysis [117]. Clinically, this disease is associated with fever, fatigue, respiratory and gastrointestinal disturbances, low levels of vascular albumin, iron, sodium and platelets, in addition to body cavity effusions amongst other symptoms [118]. This pathology is more inflammatory in nature, mostly occurring concurrently with KS. As both PEL and MCD display elevated inflammation, what makes KICS distinguishable is the higher associated tumour burden and general absence of widespread adenopathy and splenomegaly [119, 120]. KICS also appears with a CD4 count <100 cells/mm^3 as well as bone marrow related symptoms [118, 120]. KICS is the newest described KSHV-associated pathology and is difficult

to diagnose. As it stands, there is no treatment plan for KICS but rather the management of symptoms. In the largest study to date, Polizzotto *et al* reported a mortality rate of 60% and a median survival rate of 13.6 months [26].

1.5 KSHV Subtyping

The KSHV genome consists of several distinct genotypes according to the hypervariable ORF-K1 gene which is used to subtype the virus into overarching groups A, B, C, D, E and F [68]. The discovery of KSHV as well as immediate subtyping was first completed via PCR amplification and profiling using restriction enzymes [13, 50]. Today, Sanger- and next-generation sequencing (NGS) is used to determine viral subtypes based on the K1 gene [2]. However, both ORF-K1 and ORF-K15 flanking the 5' and 3' ends of the KSHV genome, respectively, show great sequence variation compared to the remainder of the genome, and both regions are currently used for KSHV subtyping. The central region of the genome gained some interest recently due to the plethora of proteins integral to viral entry and virus progression. Greater total variation observed within the central portion has created an argument to be a more reliable identifier of KSHV subtype [121]. Relatively speaking, KSHV genomes are considered highly recombinant, and differences observed in SNP density within the central region make this virus difficult to interpret and identify strain ancestry. As KSHV has been shown to evolve relatively slowly, similar to other herpesviruses [122, 123], evolutionary trends have been difficult to establish. Therefore, the K1 gene is still primarily used for subtyping purposes and can be considered the gene of interest in KSHV subtyping conventions used today.

The K1 protein is structurally similar to the lymphocyte B-cell receptor and has been found to activate cellular transformation as well as deregulating multiple B-cell signalling pathways via imitation of an activated B-cell receptor [124]. K1 has also been reported to play a role in increasing lytic replication which is vital for tumorigenesis. In the absence of the gene, results showed reduced lytic replication and decreased yields of infectious virions [68]. Variability within the K1 gene occurs as nonsynonymous mutations [50, 125, 126]. There are two distinct domains of variability within the K1 gene which are recognized as the variable regions 1 and 2 (amino acid positions 54 to 93 and 191 to 228, respectively [12]). Both regions are near 40 amino acids in length, and intratypic variability of these two regions occurs mainly within a

single 25-amino acid cysteine bridge loop [50]. ORF-K1 is approximately 289 amino acids (≈ 860 nucleotide bp) in length, with a total of 62% amino acid variation observed between subtypes A to D. Specifically, subtypes B and D differ by 30% and 24% to subtypes A and C, respectively, whilst subtypes A and C differ from each other by 15% [50]. Subtype E appears closest in similarity to subtype D; however, it differs from subtypes A to D by 25 - 30% [127]. It has been suggested that evolutionary pressure from CD8 positive cytotoxic T cells may have influenced the K1 gene's extreme variability; however, there are no known mechanisms to explain this, as the KSHV DNA polymerases cannot perform such vast changes [128, 129]. To note, individual variant clades, A1 to A10, B1 to B3 and C1 to C7 have previously been sub-grouped on the basis of unique sequence signatures within the K1 gene [50, 130-132]. KSHV evolutionary analysis conducted by Hayward and Zong [10] reports subtypes B1, B2, A5 and, more recently, subtype F in sub-Saharan Africa which still correspond to subtypes reported in the region today [2, 121, 132, 133]. Hayward and Zong hypothesized that modern humans migrated out of east Africa into sub-Saharan Africa bringing the B strain, followed by spread into south Asia, Australia and the Pacific Rim resulting in diversion into the D and E subtypes. From here, the virus appeared to migrate into the Middle East, Europe and north Asia, giving rise to the A and C strain. These observations were based on history of human migration coupled with phylogenetic branch lengths (Figure 8) [50, 125, 134]. Together, the estimated rate of change per amino acid variation is approximately 1000 years [10].

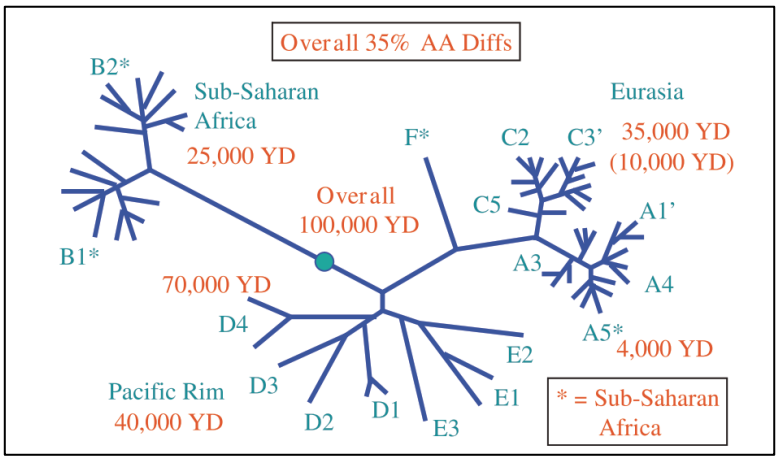


Figure 8: Radial tree depicting the KSHV K1 gene region subtypes. This phylogenetic tree illustrates the clustering, branching and relative distances between subtyped samples from Eurasia, the Pacific Rim, and sub-Saharan Africa. Image taken from Hayward and Zong, 2007 [10]. *Indicates the four uniquely sub-Saharan Africa subtypes from this study (B1, B2, A5, F).

The hypervariable K15 gene, located on the other end of the KSHV genome, produces a non-structural membrane protein [48]. It has shown to assist in angiogenesis through the recruitment and activation of PLC-gamma-1 in the transcription of pro-angiogenic genes [135] amongst other roles. K15 contains up to 33% amino acid identity and 50% similarity between the three subtypes, also referred to as alleles, which are classified as predominant (P), minor (M) and N [10, 51, 136, 137]. The P and M subtypes are the most variable, and N is considered divergent in comparison, although certain signalling motifs are conserved in all three K15 subtypes [6, 51, 136, 138]. It has been found that the K15 P subtype activates an alternative NFκβ signalling pathway and thus functions differently to the other two alleles [6, 139]. Unlike K1, K15 contains eight exons and can produce different isoforms as a result of alternative splicing. It can be hypothesized that perhaps variations within different KSHV genes can play a role in the degree of tumour development and cancer outcome.

While both the K1 and K15 genes are highly variable and have different subtype classifications, the K1 gene shows a greater association to geographical regions [6]. Therefore, K1 is to date the predominant gene for subtyping classification. It is also important to note, that the KSHV subtype does not appear to alter within an individual over time and that the subtype remains the same between tumour sites [19, 121, 140]. Variation seen today is presumed a result of slow viral evolution and intercontinental transmission over a long period of time. Subtype prevalence was found to be correlated to geographic variation, with KSHV migration over many years from surrounding regions commencing from the African continent [10]. By characterizing associated subtypes per region, one can gain invaluable insight into not only the virus' diversity and evolution but also its potential correlations to disease outcome. The KSHV subtype commonly observed in Northern Europe and America is A, whilst the B subtype is observed within the African continent and is thought to be the oldest existing strain. The C subtype is often found with classical KS cases and is typically seen in the Mediterranean, Middle East and Asia, but has since extended to Europe and the USA. Rare subtype D has mainly been described in patients from Taiwan and the Pacific Islands [50, 130, 141] and rare subtype E in Brazilian Amerindians [127] (Figure 9).

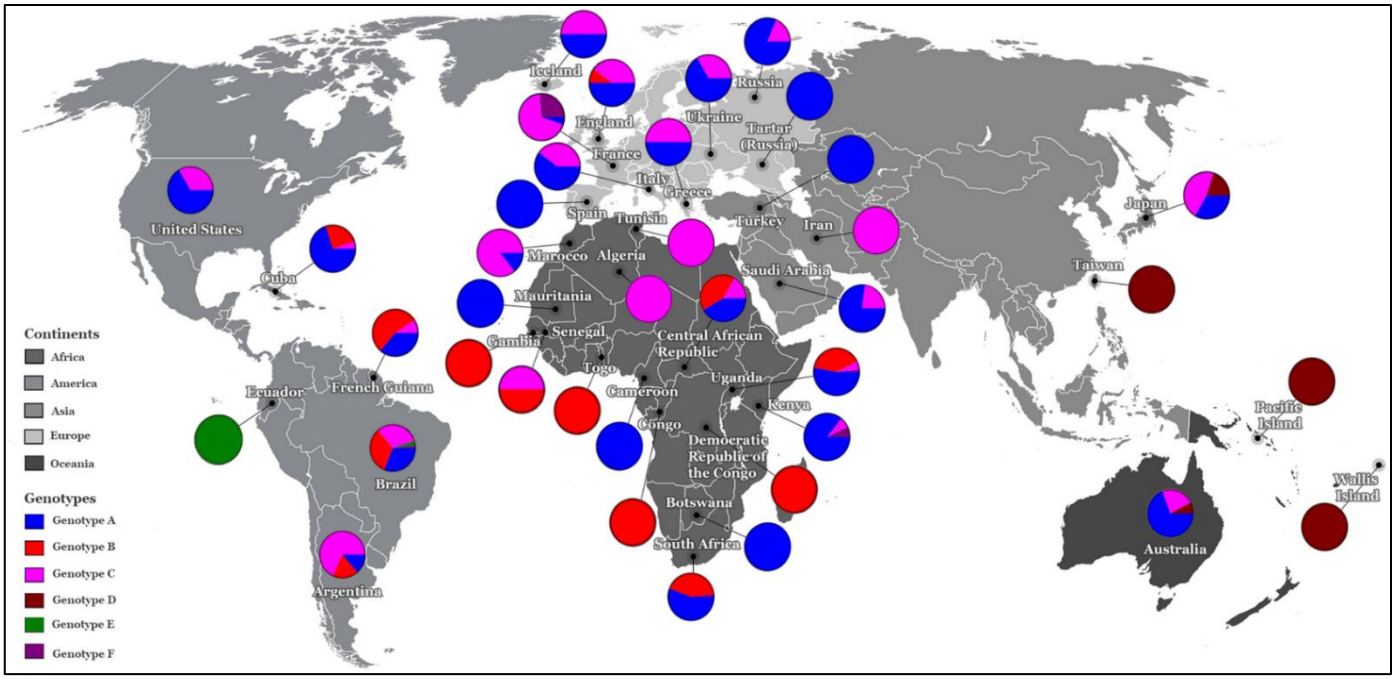


Figure 9: Worldwide map indicating global distribution of KSHV subtypes. This map provides a general indication of subtypes determined per geographical region. It is hypothesized that subtype diversity corresponds to human migration over time. Typical subtypes observed on the African continent are genotypes/subtypes A, B and C. Image taken from Lopes *et al* [3].

Subtype F was first described within the Bantu population of Uganda [132], although reports of subtypes A and B have been reported to predominate in the country [6, 131]. Figure 11 below provides an example of cluster diversity observed for different subtypes, namely those in Cameroon, West Africa. A study by Marshall *et al* documented predominant K1 subtypes A5 and B which were previously reported in this region by Betsem *et al* [2, 142]. All three K15 subtypes were detected, with the majority being the predominant P subtype.

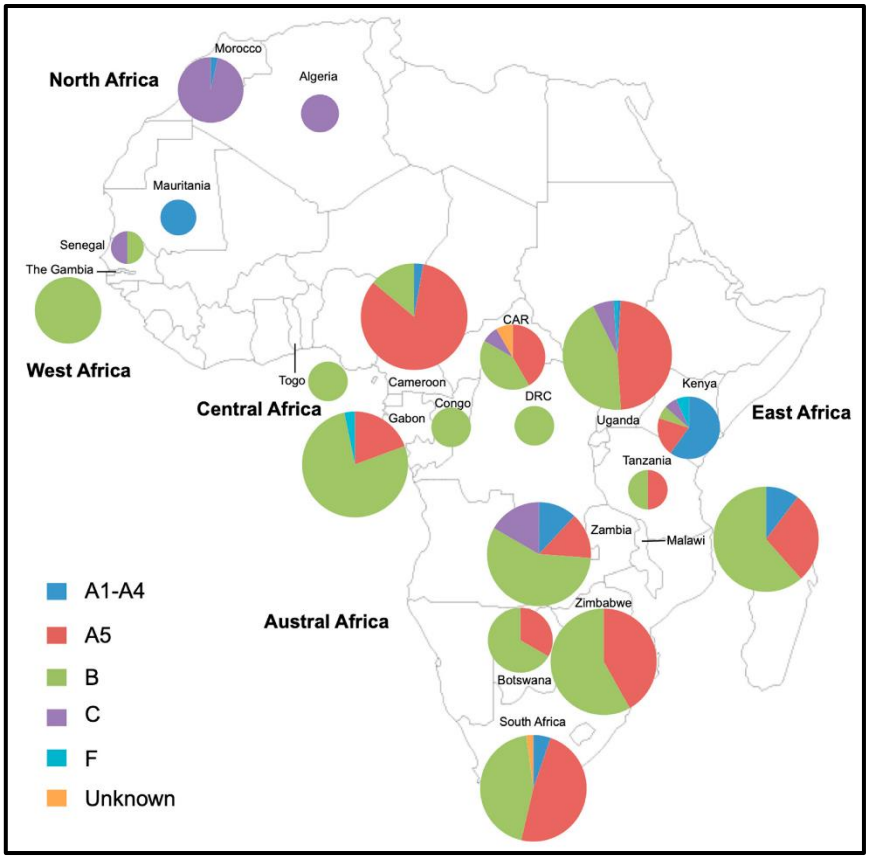


Figure 10: **Map of the African continent indicating the distribution of KSHV subtypes within various countries.** The above map has been constructed to indicate the various KSHV strains (A-F) which have been found within African countries. The sizes of the circles are proportional to the number of samples subtyped. The smallest being less than 5, intermediate is 6-30 and the largest indicates 31 or more. Image taken from Mamimandjiami et al [4].

It is important to note that a certain subtype is not unique per country, but rather there is typically a dominant KSHV subtype circulating together with other relatively minor strains per country. For example, circulating strains in Zambia were found to be A5, B, C and F with the B subtype being the most dominant strain, making up over 50% (Figure 10). This understanding is used to predict that the majority of those infected are of the B subtype.

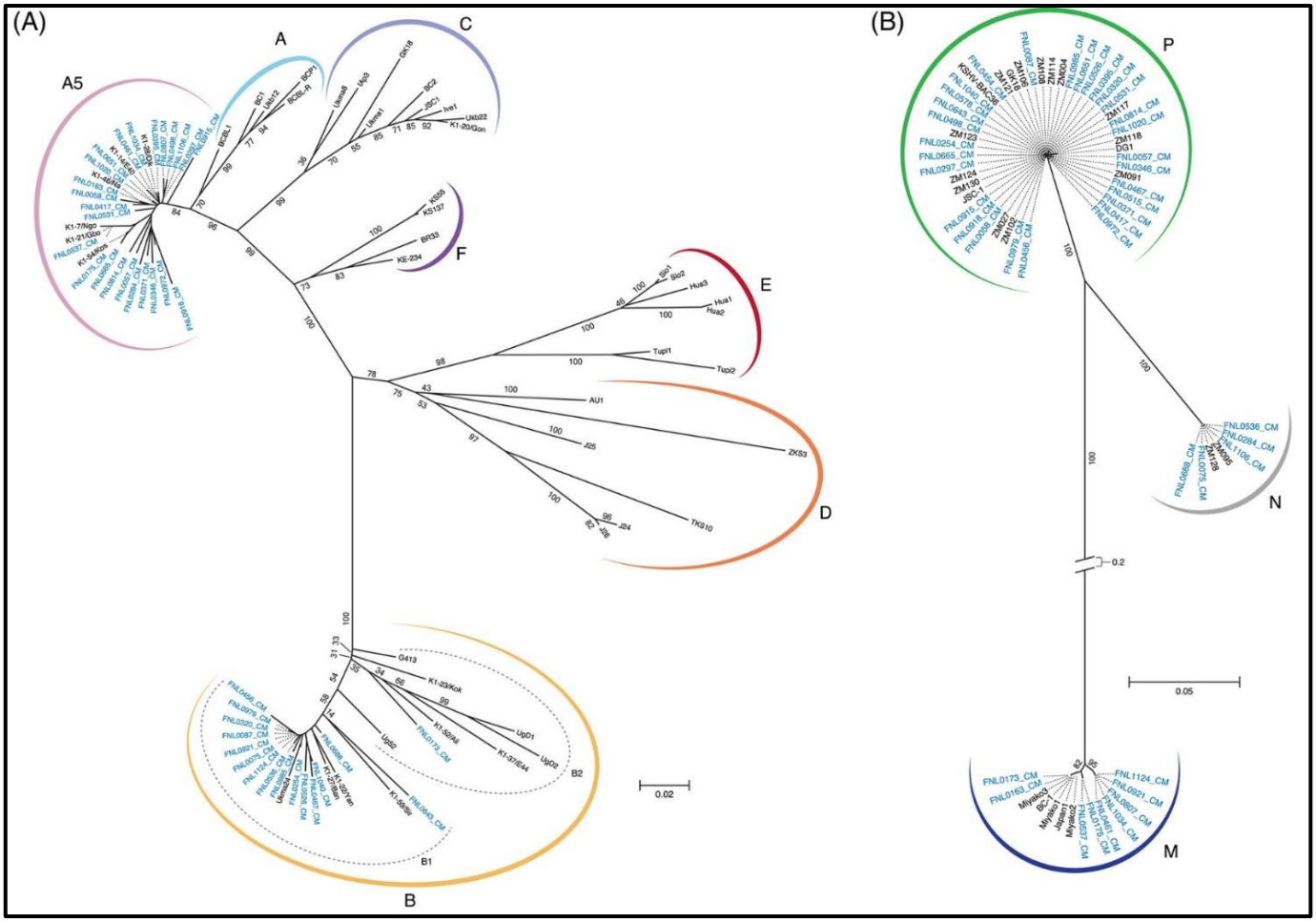


Figure 11: Examples of phylogenetic trees showing cluster diversity between KSHV subtypes. A) K1 subtypes A, A5, B, C, D, E, and F. B) K15 tree indicating subtypes M, N and P. Samples indicated in blue were sequenced from Cameroonian individuals in a study conducted by Marshall *et al* [2, 3]. Reference sequences, indicated in black, include an assortment from different geographical regions worldwide, such as Japan, USA and Uganda, some of which can be found in Table 3 in Results.

Besides using KSHV subtypes to determine historical migration of the virus and its evolutionary trends, it has been proposed that different KSHV subtypes may produce different malignant outcomes. A study of MSM living in France found that the C3 variant was associated with less severe AIDS-KS, however those with subtype A presented with higher KSHV viral loads and more severe KS [143]. Similar results for the A subtype were seen in a study by Mancuso *et al* evaluating a group of 38 Italian classic KS patients who also presented with greater disease progression [144]. A South African study found an association of increased number of lesions for the A5 genotype, lowered risk of tumour extension for A1 and A4 subtypes, and

lower likelihood of lower limb involvement for the A1 subtype [133]. All of these studies however, relied on very few patients and the interpretation of the findings should be viewed with caution.

In regard to KSHV seroprevalence testing, commonly used testing protocols include immunofluorescence assays using PEL cell lines and antigen-based ELISAs, both of which focus on reactivity to K8.1 (a lytic gene) and ORF73 (a latent gene) [145-147]. However, with these methods, seropositivity is limited to the detection of a singular antigen or both, and seroreactivity may fluctuate over time giving rise to false negatives. In addition, changes to immunity due to the presence of other viral infections, such as HIV, can influence results [148]. Broader and multiplex serodetection assays are being explored utilising the multiple KSHV-specific proteins [149]; these approaches are slowly being integrated into testing procedures [148, 150].

1.6 KSHV in Sub-Saharan Africa

KSHV is endemic to SSA and appears in parallel to KS incidence, with KS being a highly present cancer, particularly in HIV co-infected individuals [151-153]. The main mode of KSHV transmission in SSA is thought to be via saliva being passed down from mother-to-child [19, 154] likely through close contact and food sharing. This correlates with high childhood infection being observed prior to puberty [155, 156]. The percentage of KSHV infected children has previously been reported as 31.4%, 49.5%, 11.4%, 20.7% and 43.3% in Southern, Central, West, East and North Africa, respectively [157]. Seroprevalence differs per region and population, with > 50% reported in Uganda and lower prevalence being recorded in South Africa and Zimbabwe where infection may occur in early adulthood and likely through sexual practices. Uganda, Tanzania and the Congo fall inside an equatorial area often referred to as the "KS belt" where high KS incidence is observed [158]. A study conducted in Zambian infants under 12 months found a seroconversion of 50% within an average of 10.5 months [159]. More studies have noted the high seroprevalence among children in the region [17, 160-162]. Similarly, in Uganda, Butler *et al* reported a high seroprevalence of KSHV in children, with infection doubling by the age of 9 years. In adults, the seroprevalence for those 50 years and older was 49.3% [163]. To mention, increasing seroprevalence in certain regions with increased age is the result of increased sexual transmission [164] in addition to the passing of

the virus through saliva. Seroprevalence appears to be even higher in Cameroon. A study conducted using Cameroon patient samples concluded a KSHV seropositivity result of 61.5% with both HIV negative and HIV positive individuals displaying equally high KSHV antibodies [165]. Two other studies reported a seropositivity of 90% and 80% in their study participants from Cameroon [166, 167], the latter of which reported non-use of mosquito nets, treatment from a traditional healer, transfusion and family history as incidences of infection. This may shed light as to why seroprevalence is elevated within Cameroon and neighbouring regions.

In South Africa, a seroprevalence of 30.7% was recorded for the Cape Town area in 2019 [85]. A further study of South African patients from the Kwa Zulu-Natal province reported a 34.6% seropositive result in 1999 [168]. Another observed a 32% seropositive result within the Gauteng province in the same year [169]. All of which were patients above the age of 15 years, admitted to local hospitals within the area for other reasons. Malope-Kgokong *et al* reported a KSHV prevalence of 45% in women attending antenatal clinics in Gauteng province, South Africa in 2010 [18]. Although South Africa does not have as high a prevalence of KSHV as countries within and surrounding the KS belt, seroprevalence assays used may be limited or vary between studies.

K1 subtypes having been reported on the African continent include, A, A5, B, C and F with 42% being reported as B and 27% being reported as A5 (Figure 9). Subtypes A5 and B are equally dominant in a number of countries, namely South Africa [133], Botswana [170], Zambia [6], Uganda [121, 171], Central African Republic [172] and Cameroon [2] (Figure 10). For K15 subtypes, M, N and P subtypes have all been reported, with the P subtype being the most prevalent. However, only 9% of the reported K15 subtypes are of the N subtype making it the rarer genotype [157]. The K15 N subtype has been reported in South Africa and Zambia [6, 173]. In South Africa specifically, very little sequence data is available for KSHV; and to our knowledge no reported whole-genome KSHV sequences are available.

1.7 Sequencing of KSHV

In the late 90s, the beginning of KSHV research involved detection of the virus through serological assays and the establishment of cell lines from various tumour biopsies for research purposes [24, 174, 175]. Initially, “sequencing” of the virus involved a PCR amplification step followed by restriction enzyme digests to acquire specific fragments. Others used a PCR amplification and hybridization approach. After the discovery of the virus in 1994 by Chang *et al* [13], more researchers put their focus on sequencing this virus.

The first nucleotide sequence of KSHV was provided by Russo *et al* in 1996 by mapping cosmid and phage libraries generated from shot-gun sequencing to a BC-1 cell line [29]. The BC-1 cell line was established by Cesarman *et al* using lymphoma neoplasms from a patient with PEL [24]. Around the same time, the first partial KSHV sequence was attained from a classic KS biopsy originating from Greece (described by Glenn *et al* [136]). This method made use of the traditional 'one lane, four tags' Sanger approach of base calling using a traditional DNA fluorescent sequencer and today is commonly known as GK18 (NC_009333.1). The first near full-length sequence identification for KSHV was determined by Rezaee *et al* in 2006 through the assistance of the partial GK18 sequence [46]. The first group to attempt whole-genome sequencing of KSHV derived from peripheral blood and plasma using Illumina technology was Tamburro *et al* in 2012 [176]. Here, they identified a sequence called DG1 (JQ619843.1) from an HIV positive, KS negative individual when investigating KICS in a patient co-infected with both KSHV and EBV. Other well-known and frequently used KSHV sequences include BCBL-1 [177, 178] (HQ404500.1; MZ712172.1), BCBL-R [130] (AF133038), JSC-1 [179, 180] (GQ994935.1) and BC2 [130](AF133042).

Olp *et al* (2015) were the first to provide near full-length genome sequences from SSA. Using an enrichment protocol, they were able to increase coverage levels up to approximately 25 000-fold and produce 16 unique KSHV sequences from Zambia (Figure 12). Clustering was observed when running phylogenetic analysis on central portions of the genome against Western sequences, and with inclusion of highly variable K1 and K15 genes, a distinct separation was observed [6].

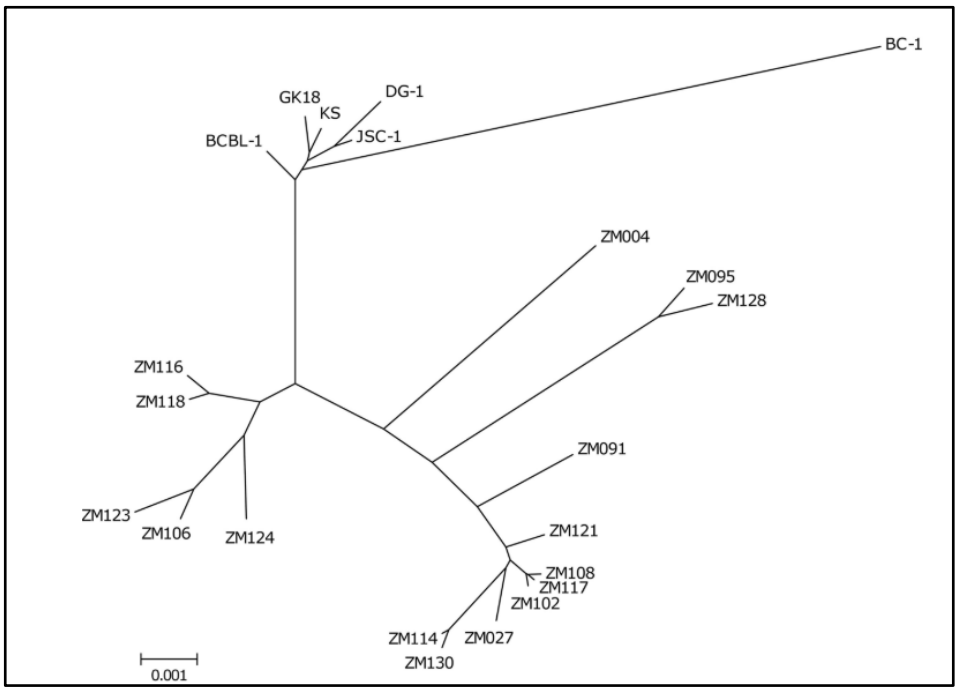


Figure 12: Example phylogenetic tree showing cluster segregation of 16 Zambian samples against 6 previously published Western samples GK18, KS, BCBL-1, DG-1, JSC-1 and BC-1. This unrooted nucleotide maximum likelihood phylogenetic tree shows KSHV sequences obtained through Illumina sequencing of tumour biopsies from Olp *et al* [6].

In South Africa, very few KSHV sequences have been submitted to GenBank to date. Those available are only derived from specific genes within the KSHV genome, such as K1 and K15. In 1997, Engelbrecht *et al* used 6 classic KS, 3 epidemic KS and 3 iatrogenic KS tumour samples from South Africa to explore KSHV variation to those reported by Zong *et al* in the same year, using the same primers [181, 182]. It was reported that minor variation was observed in the sequenced fragments, but no sequences were published. Later, in 2000, Alagiozoglou *et al* [173] successfully sequenced KSHV from South African samples using a Sanger sequencing approach covering the region of ORF75 and K15. However, only 6% of the ORF75 gene was successfully sequenced according to previous KSHV classifications of the ORF75 gene region

by Zong *et al*. The sequences were made available on GenBank (AF243797 - AF243837). Two years later, Treuricht *et al* published their findings of K1 and ORF26 sequencing results taken from 14 iatrogenic KS patients in the Western Cape, South Africa [183]. They reported on K1 A subtypes, namely A5, as well as B subtypes, namely B1 and B2. These sequences were deposited into GenBank (AF387367 - AF387378). Lastly and most recently, K1 sequences were reported from Isaacs *et al* in 2016 [133]. This study reported the largest KSHV genomic study group of the region, producing K1 subtypes for 81 AIDS-KS patients as well as 5 African endemic KS patients. Half of the study group identified as A5 (48.8%) and a near fifth as B2 18.6%), with the remainder consisting of B1, B3, A1 and A4 subtypes (32.6%). Sequences were submitted to GenBank (KP997035 - KP997135). To date, no known KSHV near full-length sequences have been reported from South Africa, making this study the first of its kind.

2. Aims and Objectives

2.1 To select +/- 50 patient DNA samples from existing patient cohorts according to specific criteria, such as minimal viral load, concentration and volume for sequencing.

2.2 To sequence the ORF-K1 gene of all DNA samples via Sanger Sequencing.

2.3 To sequence the full KSHV genome of suitable samples via Illumina NGS sequencing.

2.4 To perform sequence alignment followed by subtyping and generation of phylogenetic trees.

3. Materials & Methods

3.1 Selection of samples

The DNA samples selected for this study were retrieved from remaining material from previous studies for which ethics approval has been obtained by the University of Cape Town (UCT) Human Research Ethics Committee (HREC) (see 3.1.1, 3.1.2. and 3.1.3). The DNA had been isolated from peripheral blood (after removal of plasma) using the QIAamp DNA Blood Midi kit (Qiagen) according to the manufacturer's instructions and was provided for this study. Selection criteria were minimal detectable KSHV viral load (>1000 copies per 10 µL) as well as remaining sample volume of >5 µL (which still qualified for Sanger sequencing). Samples of greater volume and higher viral load also qualified for next generation sequencing. Table 8 (supplementary material) lists all selected samples with a viral load greater than 5 copies per 10 µL. All samples were sent to the Whitby laboratory (Frederick National Laboratory for Cancer Research, NIH, USA) where sequencing was carried out.

3.1.1 The KDHTB Study Cohort

The KDHTB (Khayelitsha Day Hospital TB) study cohort (n = 682) consisted of HIV-positive adult individuals who presented with symptoms of pulmonary or extrapulmonary TB to Khayelitsha Day Hospital, Cape Town, South Africa, between January 2014 and October 2016 in reference to the study "Defining Interventions to Reduce Mortality in Severe HIV-associated Tuberculosis" (UCT HREC/Ref: 057/2013). Eligible patients were enrolled after giving written consent, and if unconscious were enrolled and followed up daily until able to consent. Routine clinical and physical evaluations were given by physicians. Citrated blood and plasma were gathered and stored at –80°C for KSHV VL and immunologic assays.

A study published by Blumenthal *et al* [85] investigated this sample group to assess whether KSHV VL contributed to mortality in patients. Of all the samples analysed, 207 were KSHV seropositive and of this, 39 presented with a detectable KSHV VL. 20 of these 39 samples met the selection criteria for being included in the present study (Table 8 in supplementary material).

3.1.2 The KS Study Cohort

The KS study cohort consists of HIV-positive adult participants (n = 102) with a confirmed diagnosis of KS who were recruited at the Radiation Oncology Unit at Groote Schuur Hospital, Cape Town, South Africa, between December 2014 and March 2018. Informed consent was given, and ethics approval was obtained from the HREC, Health Sciences Faculty, UCT (HREC/Ref: 136/2013 and 729/2014). Citrated blood samples were collected.

These KS positive patients formed part of a larger study cohort published by Blumenthal *et al* [184] exploring EphA2 sequence variants and associated susceptibility to KSHV infection and/or KS development. This study also included samples from the above-mentioned KS-negative KDHTB cohort (3.1.1) as well as KS negative control samples received from the Infectious Disease Unit at Groote Schuur Hospital, UCT. A total of 33 patient samples from the KS patient cohort as well as one KS negative control sample (MB033) were selected to be included in this sample cohort (Table 8 in supplementary material).

3.1.3 The GUG Study Cohort

For the GUG cohort, 150 adult HIV-positive patients were recruited from the Gugulethu Community Health Centre Antiretroviral clinic (Desmond Tutu HIV Centre, UCT), South Africa, between October 2020 and June 2021. Selection criteria included a CD4 count <350 cells/µL and exclusion of TB. All participants gave informed consent and ethics approval was received from the HREC, Health Sciences Faculty, UCT (HREC Ref: 134/2020).

This patient cohort has been described in articles by Lambarey *et al* [185] exploring the association of SARS-CoV-2 infection in non-hospitalized HIV-infected patients, and Lesmes-Rodríguez *et al* [186] investigating the association of previous common coronavirus exposure to COVID-19 severity. Of the 150 samples collected, 74 reported as seropositive of which 5 presented with a detectable KSHV viral load. A total of 3 samples which met the selection criteria for this study were selected from this cohort (Table 8 in supplementary material).

3.2 ORF-K1 Sanger Sequencing

Sanger sequencing, an efficient and simple approach to sequencing, was used to sequence the K1 region of all samples to ascertain the respective KSHV subtype as this variable gene region has been widely accepted to be suitable for KSHV subtyping as mentioned in sections 1.5 and 1.7.

3.2.1 Nested PCR

Primers designed by Cook *et al* [187] were used to produce approximately 850 bp K1 amplicons using a nested PCR approach. Briefly, a PCR supermix was made up according to the number of samples in the PCR plate with extra to substantiate any pipetting error. ORF-K1 inner and outer primers (forward and reverse each, Table 1), synthesized by Eurofins Genetics, were used to create two supermixes, inner and outer respectively, for each PCR reaction. REDTaq Readymix (Sigma-Aldrich) was used in each mix at a volume of 25 µL per reaction, in addition to 0.7 µL of each respective primer, forward and reverse. 18.6 µL Nuclease-free Water (NFW) was added to achieve a total volume of 45 µL. Finally, 5 µL DNA sample was added to 45 µL outer PCR supermix within a PCR plate. The plate was placed on a Veriti 96-well thermal cycler on the program K1-Outer (95°C for 1 min, 45 sec; 35 cycles of 96°C for 1 min, 51°C for 45 sec, 72°C for 1 min; 72°C for 5 min, 4°C hold). On completion, 5 µL reaction product was added to 45 µL inner PCR supermix for each sample on a fresh plate and placed back on the PCR machine. The inner plate was run on the PCR program K1-Inner (95°C for 1 min, 45 sec; 30 cycles of 96°C for 1 min, 58°C for 45 sec, 72°C for 1 min; 72°C 5 min; 4°C hold).

Table 1: Table indicating primers used in K1 Sanger sequencing. The below table provides the inner and outer primers for forward and reverse Sanger sequencing used by Cook et al. Band size in an approximate length and may differ slightly between subtypes. Fragment size referenced from Whitby et al [170].

Gene	Primer		Size (bp)	Sequence	Reference
K1	Outer	*K1a-f*	868	ATGTTCCTGTATGTTGTCTGC	Cook et al. (1999) [187]
		K1a-r		AGTACCAATCCACTGGTTGCG	
	Inner	*K1b-f*	840	GTCTGCAGTCTGGCGGTTTGC	
		K1b-r		CTGGTTGCGTATAGTCTTCCG	

3.2.2. Gel electrophoresis and extraction

PCR K1 products from 3.2.1 were then separated on an agarose gel by electrophoresis to ascertain which samples successfully produced K1 sequence products of the expected band size. A volume of 30 µL per sample was added to a 1.5% agarose gel (SeaChem Agarose Gel, 1X TAE buffer (Tris-acetate-EDTA 50X buffer, Thermo-Fisher), 15 µL UltraPure^{TM} Ethidium Bromide, Thermo-Fisher) alongside 10 µL 1 kb ladder (1 kb TrackIt ladder, Invitrogen). 1X TAE was used as a running buffer and the gel was allowed to run at 150 V until the PCR products reached approximately two-thirds of the gel. The previous addition of REDTaq (see 3.2.1) allowed for sample visualization, while ultraviolet light was used to view the appropriate bands of the correct size which were excised using a scalpel.

3.2.3 DNA purification from agarose gels

QIAquick PCR Purification kit (Qiagen) was used to purify extracted PCR amplicons excised from the agarose gel (see 3.2.2). 500 µL of Buffer PB (provided in kit), was added to each sample and left at 4°C overnight. The samples were then applied to a QIAquick column within a collection tube and centrifuged for 60 s. The flow through was discarded and 750 µL of wash buffer PE (provided in the kit) was added and centrifuged for another 60 s. After discarding the flow through, the sample was centrifuged for another 60 s to discard any residual buffer from the column. The column was then placed in a clean 1.5 mL microcentrifuge tube and the DNA eluted using 50 µL Buffer EB (10 mM Tris-HCl, pH 8.5) and centrifuged for 1 min. Eluted DNA was measured on a Nanodrop instrument (Thermo Scientific) to confirm concentration.

3.2.4 Sanger sequencing

A Sanger sequencing reaction using the BigDye Terminator v3.1 Cycle Sequencing kit (Thermo-Fisher) was prepared individually in each well of a fresh Fast MicroAmp sequencing plate (Applied Biosystems). The reaction mix per well consisted of 8 µL BigDye Terminator, 2 µL primer (same primers used in 3.2.1, K1b-f and K1b-r, Table 1), 20 – 30 ng DNA sample (see 3.2.3) in a total volume of 20 µL which was adjusted using nuclease-free water (NFW). The plate was sealed and run on a Veriti thermal cycler on the BigDye program (96°C for 1 min; 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 4 min; 4°C hold).

The amplified DNA product was then purified to remove unincorporated salts and BigDye terminators using the BigDye XTerminator Purification kit (Thermo-Fisher). Per PCR sample, a premix was prepared using 20 µL BigDye XTerminator solution and 90 µL SAM solution. The PCR plate was centrifuged at 1000 x g for 1 min, the BigDye premix was inverted 10 times to mix and 110 µL of the premix was added to each sample well on the reaction plate. The plate was sealed using a MicroAmp™ Clear Adhesive Film, secured into a plate vortex and shaken for 20 min.

On completion, the plate was centrifuged for 2 min at 1000 xg, the plate adhesive film removed, and a capillary septum inserted on the plate. The reaction plate was then run on an ABI PRISM 3130XL DNA sequencer (3130 XL genetic analyser, Applied Biosystems/Thermo-Fisher) set to the BigDye Terminator v3.1 dye (Thermo-Fisher), Pop 7 polymer (Thermo-Fisher) and 50 cm size capillary (Thermo-Fisher). Output data was obtained from both forward and reverse sequences for all PCR products. Samples were Sanger sequenced more than once for quality control.

3.3 Whole-genome NGS library preparation and sequencing

Agilent Technologies' SureSelect^{XT} Target Enrichment System protocol for the Illumina Platform was used to prepare selected DNA samples for next-generation sequencing on the Illumina MiSeq (samples KDHTB216, KDHTB658, KS041) and Illumina NextSeq 550 (samples KS029, KDHTB153, KDHTB268, KDHTB008, MB033, MT001). Both the 200 ng and 3 µg procedures as outlined in the protocol were followed (see 3.3.1 to 3.3.8), depending on each sample's viral load, DNA concentration and volume remaining. To determine suitability, KSHV concentration of the qPCR-positive samples was determined via Qubit. With this concentration, the volume of sample necessary for the 200 ng and 3 µg protocols was calculated. If the sample did not have enough volume for either protocol, it did not qualify for sequencing. The 200 ng procedure was used in samples with low sample volume and allowed for less sample to be used. SureSelect XT Library Prep Kit (96 reactions), SureSelect Target Enrichment Kit and ILM Indexing Hyb Module, and Herculase II Fusion DNA Polymerase kit were used. Samples could be stored at –20 °C between specific steps as indicated in the protocol.

3.3.1 DNA Shearing

DNA quality, alongside concentration, was assessed on a Qubit Fluorometer and samples with an OD 260/280 between 1.8 - 2.0 were considered of high quality and purity. The DNA was then sheared to obtain appropriate DNA fragment size for sequencing as well as to produce reliable NGS results. The Covaris ML230 Focused-ultrasonicator was set up to run a programme with a 10% Duty Factor, 175 Peak Incident Power, 200 cycles per burst, a treatment time of 360 s and a bath temperature of 4°C - 8°C. 3 µg high-quality DNA was diluted with 1X Low TE Buffer (10mM Tris-HCl, pH 8.0, 0.1 mM EDTA) in a 1.5 mL LoBind tube to a final volume of 50 µL. The samples were carefully transferred through the pre-split septum of the Covaris Microtube stationed within the Covaris. The DNA was then sheared using the above-mentioned program. Samples were subsequently removed and placed into a 96-well PCR plate and topped up to 130 µL with 1X Low TE buffer.

3.3.2 Sample purification using AMPure XP beads *(included in 3 µg protocol only)*

The sheared DNA was purified to remove any fragments of unwanted size produced from the shearing step above (3.3.1). Room temperature AMPure XP beads were mixed until homogenous and 180 µL added to the sheared DNA within the PCR plate. The sample was thoroughly mixed 10 times followed by incubation for 5 min at room temperature. The plate was placed into a magnetic separation device and the cleared solution removed from each well after approximately 5 min without disturbing the beads. 200 µL of freshly prepared 70% ethanol was added to each well, mixed and subsequently removed after 1 min incubation on the magnet. This washing step was repeated. After sealing the wells, the plate was briefly spun to gather residual ethanol and returned to the magnetic stand for 30 s before removing the excess ethanol. The sample plate was then placed unsealed on a thermal cycler at 37 °C for 3 - 5 min to allow for ethanol evaporation. 50 µL NFW was added to each well, the plate sealed, mixed briefly on the vortex, and spun to collect liquid. The sample was incubated for 2 min at room temperature and then placed on a magnetic stand for 2 - 3 min until clear. The cleared super-

natant was transferred to fresh wells on the same plate (if number of samples allowed). Previously used wells on the sample plate were sealed with a strip to prevent contamination of old beads in purified samples.

3.3.3 DNA fragment quality assessment

Fragment quality was assessed using a 2100 Bioanalyzer and DNA 1000 Assay Kit (Agilent Technologies). The bioanalyzer was set up in accordance with the guide found within the Agilent DNA 1000 kit. The chip, samples and ladder were prepared according to the guide, using 1 µL of each sample (from 3.3.2). After chip preparation and loading, the instrument was run, and DNA fragment size obtained through an electropherogram. A peak size of 150 – 200 bp was considered of adequate quality for library preparation.

3.3.4 Repair of fragment ends

DNA fragment ends were repaired to allow for downstream adaptor ligation. An End Repair master mix was prepared using 35.2 µL NFW, 10 µL 10x End Repair Buffer, 1.6 µL dNTP mix, 1 µL T4 DNA polymerase, 2 µL Klenow DNA polymerase and 2.2 µL T4 polynucleotide kinase per reaction according to the SureSelect XT Library Prep Kit ILM. The 52 µL master mix was vortexed and added to each sample within the PCR plate (from 3.3.2) and mixed. The plate was incubated on the thermal cycler, without a heated lid, for 30 min at 20°C followed by a 4°C hold. Samples were purified as per step 3.3.2 above, adding 32 µL NFW for final DNA elution. The remaining sample was approximately 30 µL each.

3.3.5 Addition of dA-tail to 3’ end of DNA fragments

To allow for adaptor ligation, the DNA fragments underwent an A-tailing reaction. Using the SureSelect XT Library Prep Kit, a dA-tailing master mix was prepared using 11 µL NFW, 5 µL 10x Klenow Polymerase buffer, 1 µL dATP mix and 3 µL Exo(-) Klenow per sample. The dA-tailing master mix (20 µL total volume) was added to each end-repaired and purified DNA

sample (from 3.3.4) followed by thorough mixing. After capping, the plate was placed on a thermal cycler without a heated lid for 30 min at 37°C and then held at 4°C. Sample purification was repeated as per step 3.3.2, however half the amount of AMPure beads was added (90 µL) and only 15 µL NFW. Sample volume after purification was approximately 13 µL.

3.3.6 Ligation of the paired-end adaptors

Adaptors were ligated onto the DNA fragments (from 3.3.5) to allow for flow cell binding as well as primer recognition during amplification. A ligation master mix was prepared using 15.5 µL NFW, 10 µL 5x T4 DNA Ligase buffer, 10 µL SureSelect Adaptor Oligo mix and 1.5 µL T4 DNA ligase per sample (SureSelect XT Library Prep Kit ILM). After mixing well, the ligation mix (37 µL total volume) was added to each dA-tailed, purified DNA sample (from 3.3.5) within the PCR plate wells. The sample was mixed well, and the capped plate incubated on a thermal cycler without a heated lid for 15 min at 20°C and held at 4°C. Purification was repeated as per step 3.3.2, whereby 90 µL AMPure XP beads were added to 50 µL sample within the PCR plate. Similarly, 32 µL NFW was added to each sample following drying. 32 µL DNA library remained after purification.

3.3.7 Amplification of the adaptor-ligated library

To increase DNA signal strength during base calling, the DNA library was amplified prior to sequencing. A pre-capture PCR reaction mix was prepared using 21 µL NFW, 1.25 µL SureSelect primer, 1.25 µL SureSelect ILM Index Pre-Capture PCR reverse primer, 10 µL 5x Herculase II Reaction buffer, 0.5 µL 100 mM dNTP mix and 1 µL Herculase II Fusion DNA polymerase per reaction. The pre-capture PCR reaction mix (35 µL total volume) was vortexed and added to each well of a fresh PCR plate, followed by 15 µL purified DNA library (from 3.3.6). The remaining library was stored for future use, if needed. After mixing each sample, the PCR plate was run on a Pre-Capture PCR program (98°C for 2 min; 6 cycles of 98°C for 30 s, 65°C for 30 s and 72°C for 1 min; 72°C for 10 min followed by hold at 4°C).

Following amplification, the PCR products were purified as per step 3.3.2 using 90 µL AMPure XP beads and 30 µL NFW after drying. 30 µL sample library remained after purification.

3.3.8 DNA library quality and quantity assessment

DNA library quality was assessed by following step 3.3.3 above. An electropherogram was produced and a fragment peak between 225 bp to 275 bp for each sample was considered of adequate quality for downstream library modification steps. The library concentration was obtained through integrating the area under the peak.

3.3.9 Hybridization of DNA library

A probe hybridization-based enrichment approach was used to facilitate more comprehensive genotyping downstream. Each DNA library was adjusted to a concentration of 221 ng/µL of which 3.4 µL (i.e., 750 ng total DNA) was transferred to a new well of a 96-well plate. While capped and on ice, a Hybridization Buffer and SureSelect Block mix were made up as follows: The Hybridization buffer was prepared using 6.63 µL SureSelect Hyb 1, 0.27 µL SureSelect Hyb 2, 2.65 µL SureSelect Hyb 3 and 3.45 µL SureSelect Hyb 4 per reaction. The SureSelect Block mix was prepared using 2.5 µL SureSelect Indexing Block 1, 2.5 µL SureSelect Block 2 and 0.6 µL SureSelect Indexing Block 3 per reaction. Next, 5.6 µL SureSelect Block mix was added to each DNA library and mixed. The plate was capped and placed on a thermal cycler for 5 min at 95°C and held for at least 5 min at 65°C. A 1:9 RNase Block dilution was prepared accounting for 5 µL per hybridization reaction. A Probe Hybridization mix was prepared using 13 µL of the previously prepared Hybridization buffer mixture, 5 µL 10% RNase Block solution and 2 µL of SureSelect XT Human All Exon V8 probe per reaction. After mixing, 20 µL of this mixture was added to the DNA library-Block mix whilst in the thermal cycler maintaining 65°C and pipetted 10 times to combine. The plate was resealed using a thermal microplate film and incubated on the thermal cycler for 16 hours at 65°C with a heated lid at 105°C.

3.3.10 Capture of Hybridization-DNA library

SureSelect Wash Buffer 2 (SureSelect Target Enrichment Box 1) was prewarmed to 65°C in a water bath. Dynabeads MyOne Streptavidin T1 (Thermo Fisher) magnetic beads were vigorously resuspended and 50 µL added to wells of a fresh PCR plate. The beads were washed in 200 µL SureSelect Binding Buffer and separated using a magnetic separator device. Once the solution was clear, the supernatant was removed and discarded. The previous 2 steps were repeated for a total of 3 washes. The beads were then resuspended in 200 µL SureSelect Binding Buffer. 25 µL of the overnight hybridization-DNA library mixture (from 3.3.9) was transferred directly to each bead-containing well on the new plate and mixed vigorously on a 96-well plate mixer for 30 min at room temperature. The plate was briefly spun and placed on a magnetic separator to collect the beads. Once the solution was clear, the supernatant was discarded, and the DNA-bound beads resuspended in 200 µL SureSelect Wash Buffer 1. The solution was mixed and incubated for 15 min at room temperature. After briefly spinning in a centrifuge, the plate was again placed in a magnetic separator for the solution to clear and the supernatant was discarded. The beads were resuspended in 200 µL pre-warmed SureSelect Wash Buffer 2 and incubated on a thermal cycler at 65°C for 10 min. The plate was placed on a magnetic separator and the supernatant discarded from the cleared solution. This step was repeated a further 2 times for a total of 3 washes, after which 30 µL NFW was added to each sample well and mixed. The plate was kept on ice for indexing.

3.3.11 Primer Indexing and Purification

The DNA library samples were individually indexed to allow for sample discernment post sequencing. One indexing amplification reaction was assembled for each DNA library, using components in the kits mentioned above. The post-capture PCR reaction mix was prepared using 18.5 µL NFW, 10 µL 5x Herculase II Reaction buffer, 1 µL Herculase II Fusion DNA polymerase, 0.5 µL 100 mM dNTP mix and 1 µL SureSelect ILM Indexing Post Capture forward PCR primer per reaction. 31 µL PCR reaction mix was added to each well of a fresh PCR plate, followed by 5 µL of the appropriate indexing primer. Index assignments were made for each sample with different primers used to each sample sequenced in the same lane. The bead-

bound target-enriched DNA samples (from step 3.3.10) were mixed and 15 µL transferred to each respective well of the PCR plate containing the reaction mixture. The remaining bead mixture was stored for future use if necessary. The PCR reactions were mixed, placed on a thermal cycler, and run on the post-capture program (98°C for 2 min; 16 cycles of 98°C for 30 sec, 57°C for 30 sec and 72°C for 1 min; 72°C for 10 min; 4°C hold).

On completion, the PCR plate was briefly spun in a centrifuge. Room temperature AMPure XP beads were mixed until homogenous and 90 µL added to each 50 µL amplified DNA sample bead suspension in the PCR plate. After mixing thoroughly, the plate was left to incubate at room temperature for 5 min. The plate was placed on a magnetic stand, and the cleared solution discarded from each well, making sure not to disturb the beads. With the plate in the magnetic stand, 200 µL fresh 70% ethanol was added to each sample well. The plate was left for 1 min to allow any disturbed beads to settle, and the ethanol was discarded. Another 2 ethanol wash steps were repeated for a total of 3 ethanol washes. The plate was sealed with strip caps, briefly spun and the excess ethanol removed. The sample plate was placed unsealed on the thermal cycler set to 37°C for 2 min to dry, after which 30 µL NFW was added to each sample well. The plate was sealed, mixed thoroughly, and incubated for 2 min at room temperature. The plate was transferred to a magnetic stand and left for a further 2 min to allow the solution to clear. The cleared supernatant (approximately 30 µL) was transferred into a fresh well.

3.3.12 Index DNA library assessment and sample pooling

The quality of the amplified and indexed DNA library (from 3.3.11) was assessed by following step 3.3.3 as described above. An electropherogram was produced and a fragment peak between 250 bp to 350 bp for each sample was considered of adequate quality for sequencing. The library concentration was obtained through integrating the area under the peak.

Pooling was used to combine samples into one sequencing lane to allow for efficient and simple sequencing. The libraries were combined such that each index-tagged sample (from step 3.3.11) was present in equimolar amounts within the pool. Samples were pooled to a final volume of 20 µL and concentration of 4 nM. There were 6 indexes used per pool.

3.3.13 Whole-genome sequencing

The pooled DNA libraries (from step 3.3.12) were loaded onto the MiSeq and NextSeq, respectively. The libraries were prepared according to the Agilent SureSelect XT Custom Library Preparation kit (Agilent Technologies) as mentioned above. A custom bait set designed by Sallah *et al* [171] captured KSHV and was designed to include all published KSHV genomes from patients with PELs including BCBL-1, BC-1 and JSC-1 (see section 1.7). As samples were pooled using unique barcodes, different sample sequences could be distinguished in one run. The paired-end libraries were added into the MiSeq/NextSeq cartridge and loaded into the Illumina sequencer (500 cycle, V2 cartridge: Illumina, California, USA).

3.4 Sequence analysis

3.4.1 K1 Sanger sequence analysis

Sanger sequencing products consisting of forward and reverse signal trace, and base calling files for each sample were transferred into Geneious Prime software (Dotmatics, version 2023.2.1) for alignment. Successful sequences were manually inspected for accuracy, miscalls and frameshifts, after which the amino acid translation was secured and inspected. FASTA files were generated for both amino acid and nucleotide sequences for each respective sample, and identity verified using NCBI BLAST. DNA sequences were validated through alignment in Geneious Prime to KSHV K1 references previously published to GenBank, which contains all K1 subtypes, and subtyped accordingly. K1 sequences have been submitted to GenBank (accession numbers OR832722 – OR832741).

3.4.2 De novo pipeline used for next-generation sequence assembly

A total of 9 samples were successfully sequenced on the Illumina next-generation sequencing technologies. Files were produced in binary base call (BCL) format and automatically converted into FASTQ format for use in secondary data analysis software. After combining all four lanes for each sample using the concatenate command, a pair of mates files, R1 and R2, were produced for each sample. A pipeline was constructed for sequence assembly and alignment (Figure 13). Each pair-mates file was trimmed via 'bbduk' (BBTools, [188]) to remove sequencing adapters, after which post-trim quality control checks were run using FASTQC. Output files were filtered against KSHV and contaminant sequences using 'bbsplit '(BBTools, [188]) using several FASTA files containing human (hg19_b37), KSHV and PhiX (PhiX Control v3 control library, Illumina) assemblies, producing a singular FASTQ interleaved output file. The output underwent assembly through SPAdes (SPAdes-3.13.0, [189-191]) to align and merge DNA fragments in order to obtain the original sequence. A BLAST nucleotide database was created using the NCBI BLAST tool (ncbi-blast-2.9.0, [192]) and a FASTA file of KSHV references. For quality control, the contigs FASTA file produced from SPAdes were run through the NCBI BLAST database to align and compare the query sequences. The contigs with the top hits were isolated to produce a new, refined output containing a number of contigs of acceptable length and reliability. MeDuSa [193], a multi-draft-based scaffolder, produced a single consensus sequence from the acquired contigs with the highest BLAST hits. This scaffold was subsequently mapped using Minimap2 (minimap2-2.24_x64-linux, 10% sequence divergence [194, 195]) to a NC_009333.1 (GK18) [46] reference genome to establish the appropriate order of contigs and fully assemble the DNA sequence. This step was repeated with a reference genome most similar to the K1 and K15 subtypes for each sample. These included genomes ZM095 (KT271456.1, [6]), ZM130 (KTY271468.1, [6]), FNL0284 (OL829863.1, [2]), and BC1 (U75698.1, [13]), which were downloaded off the NCBI BLAST database and used for mapping purposes. To note, in some cases, the MeDuSa step was skipped due to poor quality output by the software and Minimap2 was used for consensus sequence generation, adjusting the sequence divergence to 20%. The .sam file outputs from mapping were uploaded into Geneious Prime to evaluate the outcome of sequencing and quality of obtained sequences.

```
#!/bin/sh
# Note. FASTQ data from the sequencer came off as one file per
lane, therefore 'cat' was used to join them into a single
file.
# STEP 1. Trimming of reads via bbduk.
# STEP 2. Running FASTQC on pre/post-trim read sets.
# STEP 3. Filtering reads against KSHV and contaminant
sequences via bbsplit.
# STEP 4. Assemble KSHV via SPAdes
# STEP 5. BLAST contigs against blastdb from step 1, pass
forward hits below e-value threshold.
# STEP 6. Getting the .tophits for the KSHV blast output.
# STEP 7. Linearizes both contigs.fasta and tophits.fasta and
saves it as filtered.fasta
# STEP 8a. Using medusa to scaffold the contigs in the hope of
producing a single consensus sequence. However, due to poorly
updated medusa, an extra step is required afterwards.
# STEP 8b. Using minimap2 to get a better consensus sequence.

#SBATCH --account=biosci
#SBATCH --partition=ada
#SBATCH --nodes=1 --ntasks=40
#SBATCH --time=56:00:00
#SBATCH --job-name="Job1_"Sample Name

#Modify the lines below for email alerts. Valid type values
are NONE, BEGIN, END, FAIL, REQUEUE, ALL
#SBATCH --mail-user=thmmel010@myuct.ac.za
#SBATCH --mail-type=BEGIN,END,FAIL

#PART ONE:
# Sample name replaced with a word for easy changing between
samples
panda="Sample Name"
# Raw reads
rawR1=/scratch/.../"Sample Name"_R1_all_lanes.fastq.gz
rawR2=/scratch/.../"Sample Name"_R2_all_lanes.fastq.gz
# Making of working directory to hold all files
mkdir /scratch/.../$panda;
# Defining of variables, "reference directory" and "working
directory"
# Once you define a variable, you can access it by typing
$refdir, $refdir, $whatever
# Here I use variables to define directories and input files
refdir=/home/.../ref/;
wdir=/scratch/.../$panda;
```

```
# Loading of modules
module load software/bbmap-38.96 software/FastQC-0.11.9
software/bbmap-38.96 software/SPAdes-3.13.0 software/ncbi-
blast-2.9.0;

# STEP 1. Trimming -
bbduk.sh in1=$rawR1 in2=$rawR2
out1=$wdir/trimmed$panda\_R1.fastq
out2=$wdir/trimmed$panda\_R2.fastq
ref=$refdir/adapters[1229].fa ktrim=r k=21 mink=8 hdist=1
qtrim=rl trimq=28 minlen=50 tbo tpe

# STEP 2. FastQC -
# Creating a directory to catch fastqc output
mkdir $wdir/fastqc
# fastqc on raw and trimmed reads
fastqc $rawR1 $rawR2 --outdir=$wdir/fastqc
fastqc $wdir/trimmed$panda\_R1.fastq
$wdir/trimmed$panda\_R2.fastq --outdir=$wdir/fastqc

# STEP 3. Filtering -
# Filtering of human reads, KSHV references, phix
bbsplit.sh
ref=$refdir/hg19_b37.fa,$refdir/KSHVrefs.fa,$refdir/phix.fa
in1=$wdir/trimmed$panda\_R1.fastq
in2=$wdir/trimmed$panda\_R2.fastq basename=$wdir/%_reads.fastq
outu=$wdir/novel.fastq ambiguous=best ambiguous2=best

# STEP 4.Assembly -
# Creatiion of assembly output folder
mkdir $wdir/Assembly
spades.py -t 80 -m 120 --12 $wdir/KSHVrefs_reads.fastq  -o
$wdir/KSHV_Assembly/ -k 21,33,55,77 --meta;

# STEP 5. BLAST -
makeblastdb -in $refdir/KSHVrefs.fa -dbtype nucl -parse_seqids
blastn -db $refdir/KSHVrefs.fa -query
$wdir/KSHV_Assembly/contigs.fasta -evalue 0.01 -out
$wdir/kshvblast.out -outfmt "6 qseqid sseqid pident length
qstart qend qlen evalue bitscore"

# PART TWO:
# changing directory to where sequencing output is
cd /scratch/...
# Output sample name...
lion="Sample Name"
# Create a working directory to hold all of this
```

```
mkdir /scratch/.../$lion
# Defining new directories and input files
refdir=/home/.../ref/;
wdir=/scratch/.../$lion;

# STEP 6. Top Hits - Getting the .tophits for the KSHV blast
output.
# Navigating within sample folder
cd ./.../$lion;
awk '!seen[$1]++ {print$1} [113]' kshvblast.out >
kshvtophits.out
mv kshvtophits.out ./KSHV_Assembly

# STEP 7. Filtering - Linearizing both contigs.fasta and
tophits.fasta and saving output as filtered.fasta
# navigate to KSHV_Assembly folder
cd ./KSHV_Assembly
awk '/^>/ {printf("%s%s\t",(N>0?"\n":""),$0);N++;next;}
{printf("%s",$0);} END {printf("\n");}' < contigs.fasta | grep
-w -A 1 -f kshvtophits.out | tr "\t" "\n" > "Sample
Name".fasta;

# STEP 8a Consensus sequence generation - Using medusa to
scaffold the contigs in the hope of producing a single
consensus sequence
# Due to poorly updated medusa, an extra step is required
afterwards.
# NOTE: Dont do this interactively and load necessary software
prior to medusa, namely compilers/gcc820, mpi/openmpi-4.0.1,
python/miniconda3-py39-usr, software/mummer-3.2.3 and
java/jdk-11
cd /scratch/.../medusa/
#load sofwares mentioned above
module load compilers/gcc820 software/medusa-1.6a
java -jar medusa.jar -f /home/.../ref/indivKSHV/ -i
/scratch/.../$lion/KSHV_Assembly/$lion\_filtered.fasta -v -d -
o /scratch/.../$lion/KSHV_Assembly/$lion\_medusascaffold.fasta

# STEP 8b. Using minimap2 to get a better consensus sequence.
# Download medusascaffold.fasta of each sequence and complete
this step using Minimap2 on your PC.
#open command terminal on PC and navigate to software folder
# sensitivity set to 10 or 20 depending on difficulty of
sequence generation
cd /mnt/.../minimap2-2.24_x64-linux$
/mnt/c/Users/...\minimap2-2.24_x64-linux/minimap2 -ax asm10
/mnt/.../"Reference Genome".fa /mnt/.../"Sample
Name"medusascaffold.fasta > /mnt/.../"Sample
Name"medusascaffold_"reference name"_asm"sensitivity".sam

# Now have consensus sequence to view in Geneious Prime
```

Figure 13: Script for running the next-generation sequencing pipeline for KSHV. The de novo pipeline script constructed for KSHV whole-genome sequence assembly. Lines beginning with # indicate a comment and not a bash script command.

3.4.3. Next-generation sequence analysis

Sequences were evaluated for completeness and annotated according to reference genome NC_009333.1 (GK18) [46], after which K1 and K15 subtype were determined via NCBI BLAST and confirmed by phylogenetic analysis. Once subtype was determined, a suitable reference genome could be acquired from GenBank for Minimap2 (as mentioned above in 3.4.2). Samples which skipped MeDuSa and Minimap alignment were pairwise aligned manually to the chosen reference genome within Geneious Prime in order to produce a consensus sequence. In the case of nucleotide disparities, the consensus sequence was called according to the generated sequence node and not the genome reference file. Once all samples had been run through the above pipeline (see 3.4.2), the raw output was again run through the de novo assembly pipeline developed by the Viral Oncology Section (VOS), Frederick National Laboratory (unpublished). Both consensus sequences for each sample underwent pairwise alignment and a final consensus was derived. Although there was little difference in output between the two sequencing pipelines, this step was to ensure that the highest sequence coverage of each sample was achieved. Further alignments were made using the de novo and reference-guided alignment (produced by VOS) output to allow verification of novel features and facilitate genome manual edits.

Nine near full-length genome KSHV partial sequences were attained and have been submitted to GenBank (accession numbers PP109054 – PP109062). Unfortunately, sequencing cannot cover the four repeat regions in the genome which is why complete genomes are rare. To note, all K15 sequences were additionally extracted from these nine near full-length sequences to be used in phylogenetic analysis.

3.5 KSHV phylogenetics

The finalized sequences, along with selected reference genomes representing circulating subtypes occurring per geographical region, were used to construct maximum likelihood phylogenetic trees for K1 and K15 gene regions, and an unrooted recombinant neighbour-net tree for the nine near full-length sequences. Both the NCI BLAST results and the phylogenetic tree topology were used to confirm KSHV subtype per sample.

3.5.1 K1- and K15-based phylogenetic trees

For the K1 phylogenetic tree, a multisequence alignment of the gathered sequences was created alongside a further 58 reference K1 sequences using MAFFT (v7.490, [196, 197]) add-on in Geneious Prime. MAFFT was set to auto algorithm with a 200PAM/k=2, 0.123 offset and 1.53 gap penalty. The alignment was then trimmed down to a length of 721 bp according to the shortest sample sequence, KS059, to allow for fair comparability between sequences as well as uniformity. The trimmed alignment was exported from Geneious Prime into a FASTA file and run through the University of Cape Town high performance computer using the iQtree software (version 2.1.3, [198-200]) with 1000 bootstrap replicates. Evolutionary models were determined empirically through the model finder setting (MFP). The same approach was carried out for the K15 phylogenetic tree, however, this tree consisted of 23 reference K15 sequences. The script for maximum likelihood tree construction can be found below in Figure 14. The iQtree output files were downloaded off the HPC and the resulting .treefile viewed in FigTree (v 1.4.4, [201]) for preliminary examination prior to being sent to the Scientific Publications department (Frederick National Laboratory, MD, USA) for illustration.

```
module                load                software/iqtree-2.1.3
iqtree2 -s "Nucleotide Alignment File".fasta -m MFP -bb 1000 -
alrt 1000 -nt AUTO
```

Figure 14: Script for running iQtree2. Bash script indicating command used to load iQtree software for phylogenetic tree construction. MFP, bootstrap set to 1000, alrt set to 1000 and nucleotide selection set to 'auto'.

3.5.2 NGS-based phylogenetic tree

A multisequence alignment of all near full-length genome sequences was constructed using MAFFT and trimmed to the first base position of BC1 on the 5’ end and the end of U004-o1 on the 3' end, to allow for uniformity and equal phylogenetic comparability. The 4 repeat regions, typical of the KSHV genome, were masked in reference to aligned and trimmed NC_009333.1 (GK18) at 23 086 bp to 30 259 bp; 35 163 bp to 35 796 bp; 124 107 bp to 130 356 bp and 135 022 bp to 142 246 bp. The sequencing bait set did not cover these repeat regions accurately and were henceforth masked. Masking was carried out via BEDTools using a bed file constructed in Microsoft Excel using the abovementioned co-ordinates and set to Linux format. The masked alignment was returned to Geneious Prime for review and then exported as a .nex file for placement into SplitsTree4 [202]. A neighbour-net SplitsTree was assembled indicating bootstrap and the file was forwarded to the Scientific Publications department (Frederick National Laboratory, MD, USA) for illustration. Recombination was evaluated through a Pairwise Homoplasy Index (PHI) text carried out in the SplitsTree4 software.

4. Results

The objective of this study was to ascertain the circulating subtypes within the region and to produce the first near full-length KSHV sequences from South Africa. A total of 29 samples derived from HIV-positive patients was sequenced and consisted of 12 KSHV-positive/KS-negative patients and 17 KSHV-positive/KS-positive patients. Of these, 18 were male and 11 were female, with a mean age of 40 years (64 - 22). All demographic and cohort information for the produced sequences as well as the consulted references are shown in Table 2 and Table 3, respectively.

The study proved successful with the sequencing of 20 K1 sequences in addition to 9 near full-length sequences (including K1 and K15) from the Western Cape of South Africa, as depicted in Figure 15. Both Sanger sequencing and Illumina sequencing platforms were used. Results confirm previous reports of subtypes A and B being present within the African continent. Furthermore, nine K15 sequences were extracted and analysed from the whole-genome data, thus providing novel K15 sequences for the region.

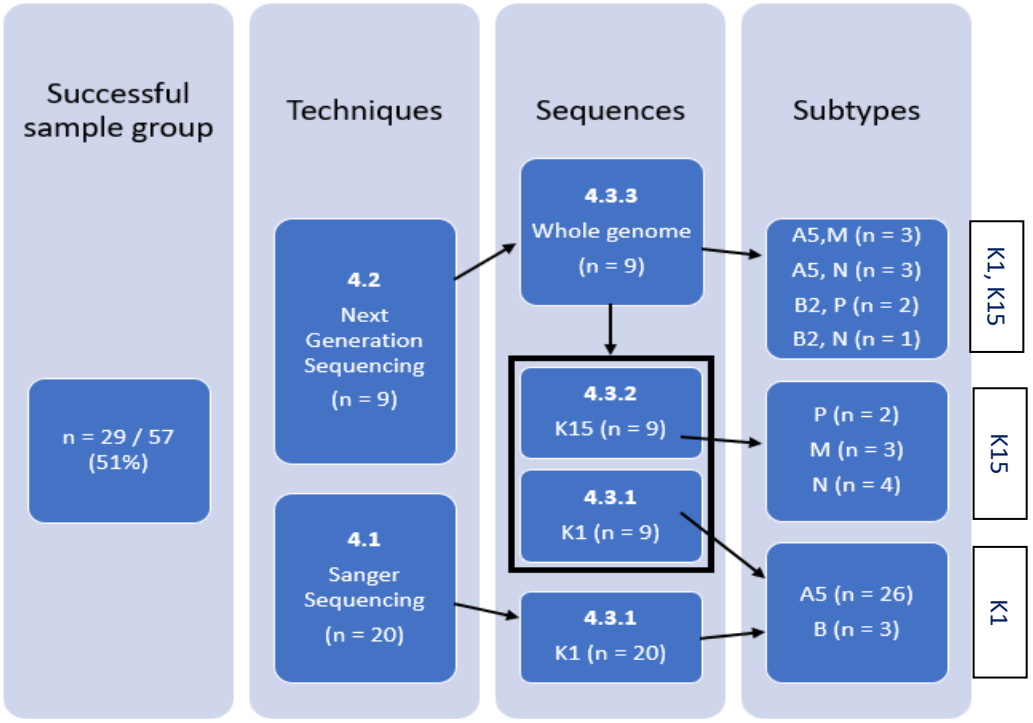


Figure 15: Flow chart providing a summary of sequencing results. This flow chart summarizes the outcome of this study. It portrays the successful sample group consisting of 51% of those originally selected. Sequencing techniques are explained in 4.1 and 4.2. Additionally, elaboration of sequencing output is explained below in 4.3.1 to 4.3.3 elaborating on the summarized subtypes. n = number of sequences.

Table 2: KSHV subtype information of the samples sequenced in this study. Summary table of all 29 successfully sequenced South African samples including demographic information as well as information on KS diagnosis and identified subtypes. GenBank accession numbers are included and will be released after publication of results. All patients were positive for HIV/AIDS. For the K15 subtypes, "-" indicate that only a K1 sequence was obtained.

#	Name	Cohort	Sex (M/F)	Age (Years)	KS (+/-)	K1	K15	GenBank accession
1	KDHTB153_SA	KDHTB	M	22	-	A5	N	PP109056
2	KDHTB658_SA	KDHTB	M	64	-	A5	N	PP109059
3	KDHTB216_SA	KDHTB	M	52	-	B2	P	PP109057
4	KDHTB268_SA	KDHTB	M	41	-	B2	P	PP109058
5	KDHTB008_SA	KDHTB	F	56	-	B2	N	PP109055
6	KDHTB566_SA	KDHTB	M	37	-	A5	-	OR832740
7	KDHTB496_SA	KDHTB	F	33	-	A5	-	OR832741
8	KDHTB296_SA	KDHTB	F	32	-	A5	-	OR832723
9	KDHTB010_SA	KDHTB	M	44	-	A5	-	OR832731
10	KDHTB096_SA	KDHTB	M	42	-	A5	-	OR832732
11	KS029_SA	KS	F	42	+	A5	M	PP109060
12	KS041_SA	KS	M	37	+	A5	N	PP109061
13	MB033_SA	KS	F	38	+	A5	M	PP109062
14	KS036_SA	KS	M	36	+	A5	-	OR832739
15	KS059_SA	KS	M	39	+	A5	-	OR832736
16	KS078_SA	KS	F	27	+	A5	-	OR832726
17	KS010_SA	KS	F	29	+	A5	-	OR832737
18	KS065_SA	KS	F	54	+	A5	-	OR832733
19	KS062_SA	KS	M	42	+	A5	-	OR832738
20	KS040_SA	KS	F	48	+	A5	-	OR832724
21	KS054_SA	KS	M	36	+	A5	-	OR832729
22	KS088_SA	KS	M	40	+	A5	-	OR832728
23	KS085_SA	KS	M	38	+	A5	-	OR832727
24	KS094_SA	KS	M	34	+	A5	-	OR832734
25	KS080_SA	KS	M	43	+	A5	-	OR832730
26	KS099_SA	KS	M	39	+	A5	-	OR832735
27	KS007_SA	KS	F	32	+	A5	-	OR832725
28	GUG116_SA	GUG	M	34	-	A5	M	PP109054
29	GUG069_SA	GUG	F	36	-	A5	-	OR832722

4.1 Sanger sequencing

In this study, 57 patient samples were selected from previously published cohorts according to the KSHV viral DNA data available (viral load, concentration and volume). Sanger sequencing focused on sequencing the K1 gene region, as per the nested PCR primers by Cook *et al* [187] and was executable at lower DNA concentrations and volume than that of next-generation sequencing. Nested PCRs as described in section 3.2.1 were performed for all samples and visualized on agarose gels. Bands occurring within the 840 bp region were extracted and purified (Figure 16). Smaller fragments occurred as a result of non-specific amplification as well as primer dimers. Four independent Sanger sequencing runs were performed on a total of 37 qualifying samples spread amongst four sequencing plates. Unsuccessful samples were repeated to ensure maximum sequence output, and selected samples were repeated to improve sequence length and quality. Those successfully producing bands in Figure 16, failed Sanger sequencing more than once, likely due to poor sample quality. Of the 37 samples sequenced, 20 samples produced successful K1 sequences. These were of acceptable quality and were confirmed by their corresponding trace files. Each consensus generated consisted of at least one complete forward and one reverse sequence per sample. These sequences were later trimmed to a standard length of 721 bp, covering both VR1 and VR2 within the gene. All 20 K1 sequences have been submitted to GenBank (accession numbers OR832722 – OR832741) and will be released shortly.

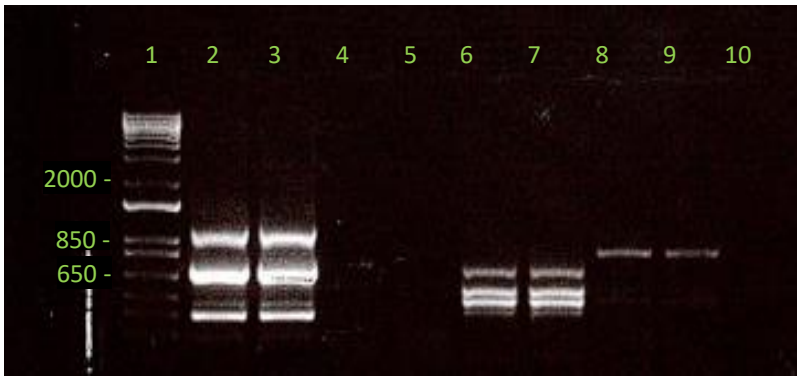


Figure 16: A 1.5% Agarose gel illustrating DNA bands post-PCR amplification. Bands sitting at approximately 850 kb were extracted under UV and purified. In this gel image, the first band was extracted from lanes 2, 3, 8 and 9, with samples occurring in duplicate. Smaller size fragments indicate the presence of non-specific binding and primer dimers. Lane 1 contains the TrackIt 1 kb ladder. Lane 2 & 3: KS039; lane 4 & 5: GUG064; lane 6 & 7: KDHTB094; lane 8 & 9: KS016, lane 10: negative control.

It is important to make mention that K15 did not undergo Sanger sequencing for the 20 samples that underwent K1 sequencing. This was due to very limited volumes of DNA remaining. The large variability between K15 alleles does not allow the designing of primers that would be able to detect all three subtypes (P, M and N) in a single reaction. Each subtype would need to be evaluated in separate nested PCR assays increasing the amount of DNA required.

4.2 Next-generation sequencing

A targeted enrichment protocol was used to prepare ten suitable samples from the available 57 individual samples, namely KDHTB153_SA, KDHTB658_SA, KDHTB216_SA, KDHTB268_SA, KDHTB008_SA, KDHTB528, KS029_SA, KS041_SA, MB033_SA and GUG116_ SA. Of these, nine produced near full length genome sequences as one sample, KDHTB528 did not sequence successfully (Table 9, supplementary material). Sequence suitability was determined via the 3 µg and 200 ng enrichment protocols as mentioned (see 3.3.). KSHV viral load >6000 copies/10 µL have previously been shown to produce good results [2]. KDHTB008 and KDHTB216 presented with a viral load of approximately 1200 raw copies/10 µL each; however, other samples with lower copy numbers were included if presenting with a promising DNA concentration and high volume. When using the NextSeq, we found the viral load could be dropped to a minimum of 1000 raw copies/10 µL which allowed for more samples to be run.

The majority of samples had viral loads well below 1000 copies/10 µL and therefore did not qualify for full genome sequencing; however, most were suitable for Sanger sequencing (see 4.1). Samples within an indexed library were loaded onto the Illumina sequencer and run for approximately 24 hours. All samples produced good coverage sequence. Of note, sample KDHTB008 was sequenced successfully with the exception of a large gap within ORF K15 (see 4.3.2). Interestingly, sample GUG116 reported a remarkably low viral load of 643 copies/ 10 µL, yet produced no errors during sequencing. Sequencer files were run through the constructed de novo assembly pipeline (see 3.4.2, Figure 13) to produce sequence alignment map (.sam) files which were examined in Geneious Prime. Sequence coverage can be found in Table 6 section 4.3.3. Additionally, the sequence files were run through a second de novo pipeline created by the VOS, Frederick National Laboratory, MD, USA, in order to ensure maximum sequence quality. Pipeline output was near identical and confirmed the quality of the pipeline

constructed above (see 3.4.2). Both .sam files generated were aligned in Geneious Prime and produced nine near full length consensus sequences of acceptable quality, namely KDHTB153_SA, KDHTB658_SA, KDHTB216_SA, KDHTB268_SA, KDHTB008_SA, KS029_SA, KS041_SA, MB033_SA and GUG116_ SA. The sample sequences were aligned and trimmed from both the 5' end prior to the start of K1 as well as the 3' end after the K15 gene. All 9 near full length KSHV sequences have been submitted to GenBank (accession numbers PP109054 – PP109062 and will be released shortly.

4.3. Subtyping and phylogenetic analysis

For phylogenetic analysis, a number of suitable reference sequences were selected. Sequences were chosen from well known, established studies ([2, 121, 143, 203]) and chosen as representatives of geographical origin. In regard to subtypes, a number of samples were selected for each subtype per phylogenetic tree (Figures 16, 18 and 20) to ensure balance and adequate portrayal of each subtype's unique signatures. All sequences were downloaded from GenBank with the exception of four samples from The European Nucleotide Archive. Details can be found in Table 3.

Table 3: Table indicating all reference sequences used in phylogenetic tree construction. Shown are the sequence names, accession numbers, country of sample origin and subtypes for all KSHV reference sequences used in phylogenetic tree construction for the K1, K15 and NGS trees. Not all available sequences are shown. The relevant publications are indicated. (*Sequence available within The European Nucleotide Archive as opposed to GenBank).

Sequence name	GenBank Accession no.	Country	K1	K15	Reference
Japan1	LC200589	Japan	C	M	Awazawa *et al* [203]
Miyako1	LC200586	Japan	C	M	
Miyako2	LC200587	Japan	C	M	
Miyako3	LC200588	Japan	C	M	
FNL0058	OL829883	Cameroon	A5	P	Marshall *et al* [2]
FNL0284	OL829863	Cameroon	A5	N	
FNL0467	OL829886	Cameroon	B1	P	
FNL0536	OL829862	Cameroon	B1	N	
FNL0814	OL829867	Cameroon	A5	P	
FNL1040	OL829885	Cameroon	B1	P	
FNL1106	OL829864	Cameroon	A5	N	
P030_KS	MK876732	France	F	P	Jary *et al* [143]

P044_PEL	MK876733	France	A4	P	
P075_MCD	MK876735	France	F	P	
P076_PEL	MK876736	France	F	P	
U004-o1	MT510664	Uganda	C1	P1	Santiago *et al* [204]
U007-B	MT510654	Uganda	B1	P1	
U008-B	MT510656	Uganda	B1	P1	
U020-B	MT510666	Uganda	A5	P2	
U023-o1	MT510669	Uganda	B3	P2	
U030-C	MT510670	Uganda	B1	M2	
U032-B	MT510652	Uganda	A5	P1	
U034-B	MT510659	Uganda	B1	P1	
BCBL1/BAC36	MT936340	USA	A3	P1	
UNC_RM_23	MZ712177	Malawi	A5	P	Moorad *et al* [205]
UNC_RM_52	MZ712182	Malawi	A5	P	
UNC_RM_98	MZ712179	Malawi	A5	P	
UNC_RM_269	MZ712178	Malawi	B3	P	
ZM004	KT271453	Zambia	B1	P	Olp *et al* [6]
ZM095	KT271456	Zambia	B4	N	
ZM121	KT271464	Zambia	B1	P	
ZM123	KT271465	Zambia	B1	P	
ZM128	KT271467	Zambia	B1	N	
ZM130	KT271468	Zambia	B3	P	
GK18	NC_009333.1	Greece	C	P	Glenn *et al* [136]
JSC1	GQ994935	USA (cell line)	C	P	Brulois *et al* [179]
BC1	U75698.1	USA	A2	M	Cesarman *et al* [24]
UG118	ERS1615777*	Uganda	B1	M	Sallah *et al* [171]
UG119	ERS1615780*	Uganda	A5	M	
UG128	ERS1615712*	Uganda	A5	P	
UG151	ERS1615793*	Uganda	B1	P	
K1-7/Ngo	AF178779	Central African Republic	A5	-	Lacoste *et al* [172]
K1-20/Gon	AF178789	Central African Republic	C	-	
K1-21/Gbo	AF178790	Central African Republic	A5	-	
K1-28/Djk	AF178797	Cameroon	A5	-	
K1-37/E44	AF178804	Togo	B2	-	
K1-52/Ali	AF178818	Senegal	B2	-	
K1-59/Sir	AF178824	Central African Republic	B1	-	
Hua1	AY329027	Ecuador	E	-	Whitby *et al* [170]
Hua2	AY329028	Ecuador	E	-	
Hua3	AY329026	Ecuador	E	-	

KE-234	FJ884616	Kenya	F	-	Tornesello *et al* [131]
KS40	KP997061	South Africa	A	-	Isaacs *et al* [133]
KS41	KP997062	South Africa	B2	-	
KS42	KP997063	South Africa	A5	-	
KS46	KP997067	South Africa	B3	-	
KS78	KP997096	South Africa	B1	-	
J24	AF278844	Japan	D	-	Meng *et al* [206]
J25	AF278845	Japan	D	-	
J26	AF278846	Japan	D	-	
Iap3	AF130271	Italy	C	-	Cook *et al* [187]
Ive1	AF130286	Italy	C	-	
K1-46/Na	AF171058	Cameroon	A5	-	Fouchard *et al* [207]
K1-54/Kos	AF171059	Central African Republic	A5	-	
Tupi-1	AF220292	Brazil	E	-	Biggar *et al* [127]
Tupi-2	AF220293	Brazil	E	-	
UgD1	AF130292	Uganda	B3	-	Cook *et al* [187]
UgD2	AF130293	Uganda	B3	-	
UKb22	AF130298	United Kingdom	C	-	
UKma1	AF130300	United Kingdom	C	-	
Ug3/ug13	AF151690	Uganda	F	-	Meng *et al* [208]
Au1/407p	AF151687	Australia	D	-	
BC2	AF133042	USA (cell line)	C	-	Nicholas *et al* [130]
BCBL-R	AF133038	USA (cell line)	A	-	
ZKS3	AF133044	New Zealand	D2	-	
Bot3	AY329023	Botswana	A5	-	Whitby *et al* [170]
N1	DQ394068.1	Morocco	C	-	Duprez *et al* [209]
K1-45/Min	AF178812	Mauritania	A2	-	Lacoste *et al* [172]
K1-43/Berr	AF178810	France	F	-	
BR33	KT215106.1	Brazil	F	-	Tozetto-Mendoza *et al* [210]
431K	MH613917	DRC	-	P	Poole *et al* [51]
APK1	MH613918	Italy	-	P	
ASM70-80	MH613961	USA	-	M	
C282	MH613923	USA	-	P	
FTKS8	MH613924	South Africa	-	P	
HKS11	MH613963	Uganda	-	M	
HKS18	MH613927	Uganda	-	P	
HKS35	MH613975	Uganda	-	N	
HKS49	MH613964	Uganda	-	M	
HKS58	MH613939	Uganda	-	P	
RKS1	MH613944	Zambia	-	P	
SAKS25	MH613960	South Africa	-	N	
SAKS28	MH613971	South Africa	-	P	Zong *et al* [125]

SAKS30	MH613972	South Africa	-	N	Poole *et al* [51]
SKS1	MH613955	Saudi Arabia	-	P	
TKS10	MH613956	Taiwan	-	P	
TKS13	MH613968	Taiwan	-	M	
ZKS3	MH613957	New Zealand	-	P	
ZKS4	MH613958	New Zealand	-	P	
Ugd2	AY043002	Uganda	-	P	Kakoola *et al* [211]
Ugd4	AY043008	Uganda	-	P	
Ugd10	AY042965	Uganda	-	M	
Ugd12	AY042966	Uganda	-	P	

4.3.1 K1 gene

Overall, this study produced 29 consensus sequences for ORF-K1 from Sanger and whole-genome sequencing collectively. K1 gene sequences were run through NCBI BLAST to confirm subtypes based on readily available and published K1 sequences. In addition, a multiple alignment with reference genomes and subsequent phylogenetic tree construction confirmed K1 subtype. Of the 29 K1 sequences, 26 (89.7%) were of the A5 subtype and the remaining three were classified as B2 (10.3%) as per the established subtyping conventions used today.

For phylogenetic tree construction, the 20 K1 South African sequences produced in this study were aligned to 59 K1 reference sequences of subtypes A through F, and trimmed to a standard length of 721 bp. These K1 references included reference sequences from various African countries such as Uganda, Cameroon and Botswana, as well as other non-African countries, such as Japan, the United Kingdom and Brazil (Table 3).

The phylogenetic tree depicted in Figure 17 below, provides not only K1 subtype confirmation of the samples included in this study, but additionally a strong geographical signal for these subtypes. The topology reveals that the A and B subtypes sit at opposite poles of the tree, thus portraying the greatest sequence diversity. All gathered K1 samples from this study, as well as others reported from South Africa [133], are found within these two branches. Twenty-six K1 coding regions, sequenced from our South African samples, fell within the A5 branch. In addition, the A5 branch includes samples from the Central African Republic (K1-54/Kos and K1-7/Ngo) Cameroon (FNL0814, FNL0058, K1-28/Dijk and K1-46/Na) as well as Uganda (Ug3, UG119, U020-B and UG128) (Table 3). Interestingly, a sample sequenced from

Botswana, a neighbouring country to South Africa, fell amongst our South African samples, suggesting close genetic ties to the Southern region of the continent. In this study, three samples sequenced were of the K1 B2 subtype, namely KDHTB216, KDHTB008 and KDHTB268, which is also a dominant subtype circulating within the continent. These samples share the B branch with two other South African K1 samples published by Isaacs *et al* [133], namely KS41 and KS46, as well as samples from Uganda (U023-o1, UgD1, UgD2, UG118) and Zambia (ZM130, ZM095, ZM121) as well as others from Senegal (K1-52/Ali), Togo (K1-37/E44) and Central African Republic (K1-59/Sir) (Table 3). The constructed tree confirms previous findings that the K1 subtypes A and B predominate within Africa [2, 171, 172, 187].

In addition to subtype confirmation, the phylogenetic tree in Figure 17 illustrates the degree of variability between African and non-African samples. Low variability can be found within the A branch, adjacent to A5, and includes the USA cell lines, BCBL1, BCBL-R and BC1. Similarly, the same phenomenon is observed within the C branch, which includes GK18 from Greece as well as Miyako samples 1 and 2 from Japan. However, it has been reported that samples originating from different countries have presented with near identical K1 sequences suggesting that geographical location may not correlate with KSHV subtype as much as previously thought [187]. Yet, infection with different geographical strains can be as a result of human travel and migration. In contrast to this, high K1 variance can be observed for branches D, E and F in relation to the South African samples situated in the A and B branch. This would include K1 sequences from France in the F branch, Hua1 through Hua3 from Ecuador in the E branch in addition to J24 to J26 from Japan in the D branch. All of the above sequences included are typical of the subtype per region.

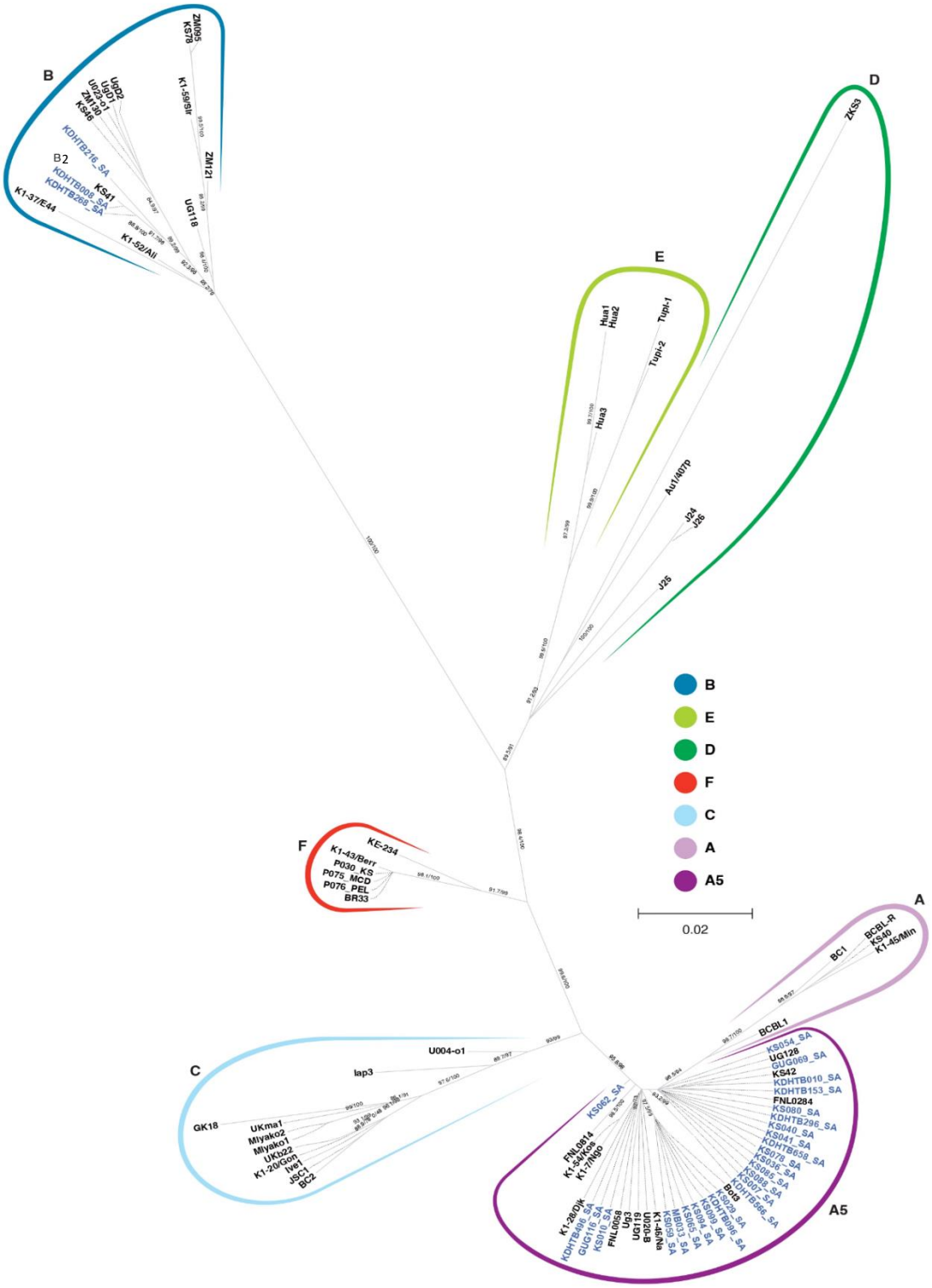
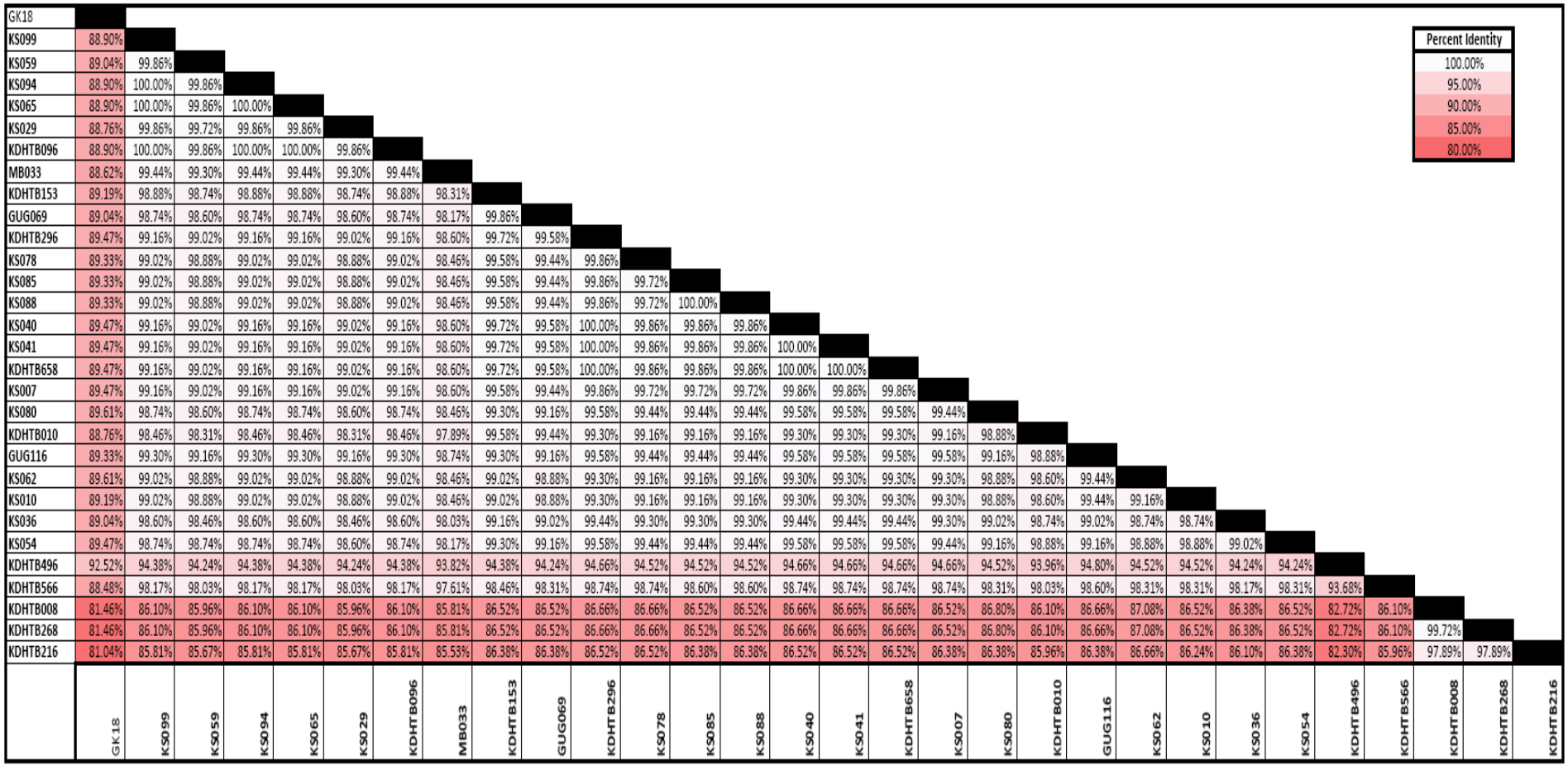


Figure 17: Maximum likelihood tree of the K1 gene region. The phylogenetic tree illustrates where the obtained K1 South African sequences (indicated in blue and by '_SA') lie in relation to other subtypes from different geographical regions worldwide. Branches correspond to subtype (A-F) and are indicated in portrayed in varying colours. This maximum likelihood tree was generated using selected reference sequences and iQtree software. Bootstrap values are indicated on each branch. Graphics are courtesy of the Scientific Publications Department (NIH/NCI).



	GK18	KS099	KS059	KS094	KS065	KS029	KDHTB096	MB033	KDHTB153	GUG069	KDHTB296	KS078	KS085	KS088	KS040	KS041	KDHTB658	KS007	KS080	KDHTB010	GUG116	KS062	KS010	KS036	KS054	KDHTB496	KDHTB566	KDHTB008	KDHTB268	KDHTB216
GK18																														
KS099	88.90%																													
KS059	89.04%	99.86%																												
KS094	88.90%	100.00%	99.86%																											
KS065	88.90%	100.00%	99.86%	100.00%																										
KS029	88.76%	99.86%	99.72%	99.86%	99.86%																									
KDHTB096	88.90%	100.00%	99.86%	100.00%	100.00%	99.86%																								
MB033	88.62%	99.44%	99.30%	99.44%	99.44%	99.30%	99.44%																							
KDHTB153	89.19%	98.88%	98.74%	98.88%	98.88%	98.74%	98.88%	98.31%																						
GUG069	89.04%	98.74%	98.60%	98.74%	98.74%	98.60%	98.74%	98.17%	99.86%																					
KDHTB296	89.47%	99.16%	99.02%	99.16%	99.16%	99.02%	99.16%	98.60%	99.72%	99.58%																				
KS078	89.33%	99.02%	98.88%	99.02%	99.02%	98.88%	99.02%	98.46%	99.58%	99.44%	99.86%																			
KS085	89.33%	99.02%	98.88%	99.02%	99.02%	98.88%	99.02%	98.46%	99.58%	99.44%	99.86%	99.72%																		
KS088	89.33%	99.02%	98.88%	99.02%	99.02%	98.88%	99.02%	98.46%	99.58%	99.44%	99.86%	99.72%	100.00%																	
KS040	89.47%	99.16%	99.02%	99.16%	99.16%	99.02%	99.16%	98.60%	99.72%	99.58%	100.00%	99.86%	99.86%	99.86%																
KS041	89.47%	99.16%	99.02%	99.16%	99.16%	99.02%	99.16%	98.60%	99.72%	99.58%	100.00%	99.86%	99.86%	99.86%	100.00%															
KDHTB658	89.47%	99.16%	99.02%	99.16%	99.16%	99.02%	99.16%	98.60%	99.72%	99.58%	100.00%	99.86%	99.86%	99.86%	100.00%	100.00%														
KS007	89.47%	99.16%	99.02%	99.16%	99.16%	99.02%	99.16%	98.60%	99.58%	99.44%	99.86%	99.72%	99.72%	99.72%	99.86%	99.86%	99.86%													
KS080	89.61%	98.74%	98.60%	98.74%	98.74%	98.60%	98.74%	98.46%	99.30%	99.16%	99.58%	99.44%	99.44%	99.44%	99.58%	99.58%	99.58%	99.44%												
KDHTB010	88.76%	98.46%	98.31%	98.46%	98.46%	98.31%	98.46%	97.89%	99.58%	99.44%	99.30%	99.16%	99.16%	99.16%	99.30%	99.30%	99.30%	99.16%	98.88%											
GUG116	89.33%	99.30%	99.16%	99.30%	99.30%	99.16%	99.30%	98.74%	99.30%	99.16%	99.58%	99.44%	99.44%	99.44%	99.58%	99.58%	99.58%	99.58%	99.16%	98.88%										
KS062	89.61%	99.02%	98.88%	99.02%	99.02%	98.88%	99.02%	98.46%	99.02%	98.88%	99.30%	99.16%	99.16%	99.16%	99.30%	99.30%	99.30%	99.30%	98.88%	98.60%	99.44%									
KS010	89.19%	99.02%	98.88%	99.02%	99.02%	98.88%	99.02%	98.46%	99.02%	98.88%	99.30%	99.16%	99.16%	99.16%	99.30%	99.30%	99.30%	99.30%	98.88%	98.60%	99.44%	99.16%								
KS036	89.04%	98.60%	98.46%	98.60%	98.60%	98.46%	98.60%	98.03%	99.16%	99.02%	99.44%	99.30%	99.30%	99.30%	99.44%	99.44%	99.44%	99.30%	99.02%	98.74%	99.02%	98.74%	98.74%							
KS054	89.47%	98.74%	98.74%	98.74%	98.74%	98.60%	98.74%	98.17%	99.30%	99.16%	99.58%	99.44%	99.44%	99.44%	99.58%	99.58%	99.58%	99.44%	99.16%	98.88%	99.16%	98.88%	98.88%	99.02%						
KDHTB496	92.52%	94.38%	94.24%	94.38%	94.38%	94.24%	94.38%	93.82%	94.38%	94.24%	94.66%	94.52%	94.52%	94.52%	94.66%	94.66%	94.66%	94.66%	94.52%	93.96%	94.80%	94.52%	94.52%	94.24%	94.24%					
KDHTB566	88.48%	98.17%	98.03%	98.17%	98.17%	98.03%	98.17%	97.61%	98.46%	98.31%	98.74%	98.74%	98.60%	98.60%	98.74%	98.74%	98.74%	98.74%	98.31%	98.03%	98.60%	98.31%	98.31%	98.17%	98.31%	93.68%				
KDHTB008	81.46%	86.10%	85.96%	86.10%	86.10%	85.96%	86.10%	85.81%	86.52%	86.52%	86.66%	86.66%	86.52%	86.52%	86.66%	86.66%	86.66%	86.52%	86.80%	86.10%	86.66%	87.08%	86.52%	86.38%	86.52%	82.72%	86.10%			
KDHTB268	81.46%	86.10%	85.96%	86.10%	86.10%	85.96%	86.10%	85.81%	86.52%	86.52%	86.66%	86.66%	86.52%	86.52%	86.66%	86.66%	86.66%	86.52%	86.80%	86.10%	86.66%	87.08%	86.52%	86.38%	86.52%	82.72%	86.10%	99.72%		
KDHTB216	81.04%	85.81%	85.67%	85.81%	85.81%	85.67%	85.81%	85.53%	86.38%	86.38%	86.52%	86.52%	86.38%	86.38%	86.52%	86.52%	86.52%	86.38%	86.38%	85.96%	86.38%	86.66%	86.24%	86.10%	86.38%	82.30%	85.96%	97.89%	97.89%	

Table 4: Distance table of the K1 coding region indicating the percent identity in reference to NC_009333.1 (GK18) for each South African sample of this study. Shown is percentage of similarity (percent identity) of the K1 sequence of each sample to that of GK18. Samples KDHTB008, KDHTB268 and KDHTB216 are of the B subtype, GK18 of the C subtype and the remainder fall under the A5 subtype for K1 classifications. Low percentage identity is indicated in red.

Assessment of the percent identity of the K1 region of our sequences compared to that of the of the reference sequence GK18 (subtype C) as shown in Table 4 confirms the subtype assignment illustrated in the phylogenetic tree (Figure 17). The lowest percent identity, or similarity, is observed for the three B subtypes, KDHTB008, KDHTB268 and KDHTB216, while the remainder A5 subtypes show substantially higher percent identity. The A5 subtypes display an average percent identity of 88.47% to GK18, and the B subtype produced an average of 81.32% percent identity to that of GK18. As GK18 is a C subtype, it is expected that A5 subtypes will have a greater percent similarity as this branch is closer to the A branch than the distant B branch, as visualised in Figure 17. To note, sample KDHTB496 presents with a 36 bp gap (5% of the CDS region) which can be seen in Figure 18 and Table 4. This 36 bp gap may be a result of poor coverage during sequencing or may indeed be a true indel (i.e., insertion or deletion) in the KSHV DNA. GK18 also presents with this phenomenon in the same region. However, as the K1 gene displays high variability and thus multiple subtypes, a portion as short as 300 bp can provide subtype classification as long as at least one of the two variable regions is accounted for. Therefore, taking this into consideration, a missing region of 5% in KDHTB496 poses no conflict for subtype classification.

To conclude, the observed percent identities follow the expected trends and additionally confirms the above-mentioned K1 subtyping.

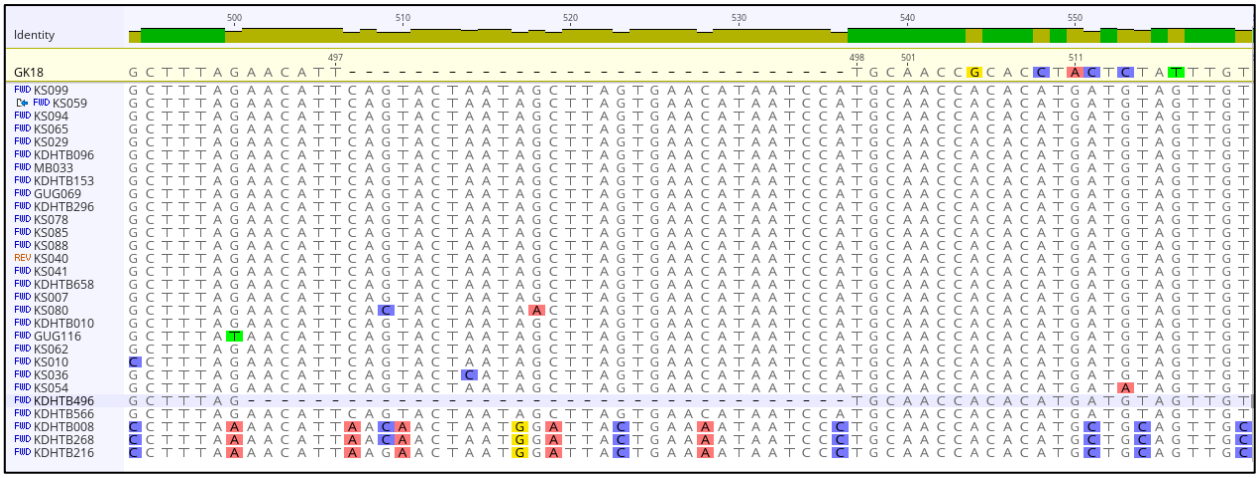


Figure 18: K1 alignment illustrating a portion of the gene whereby a 36 bp gap is observed in KDHTB496. This may be due to poor coverage or a true indel. To note, samples KDHTB008, KDHTB268 and KDHTB216 are of the K1 B2 subtype and contain similar nucleotide identities. The remainder of the sequences shown are of the A5 subtype with the exception of reference sequence GK18 which is a C subtype.

4.3.2 K15 gene

Nine K15 sequences were extracted from the nine near full-length genome sequences acquired and were subtyped accordingly. Sequences were subtyped by the use of the NCBI BLAST database and confirmed through phylogenetic tree analysis. Of the nine sequences gathered, all three K15 subtypes were discovered measuring 2186 bp in length, with 2/9 (22%) being of the P subtype, 3/9 (33%) of the M subtype and 4/9 (44%) of the rare N subtype as per the currently used subtyping conventions.

The K15 phylogenetic tree illustrates the degree of variance between the three K15 subtypes as visualized by the unique tree topology (Figure 19). The M, N and P subtypes vary greatly, with subtypes P and M displaying the greatest genetic divergence, as depicted through branch distances. Within the P branch, KDHTB268 and KDHTB216 lie close to the selected reference sequences originating from the Democratic Republic of Congo (431K), Uganda (HKS58, Ugd2, Ugd4, Ugd12), Zambia (RKS1) as well as other K15 sequences gathered from South Africa (FTKS8 and SAKS28) (Table 3). This suggests close sequence similarity for the K15 P allele, which is to be expected due to the close geographic origin. Within the M branch, samples MB033, KS029 and GUG116 are near identical to the K15 of ASM70-80 from the USA followed by TKS13 from Taiwan, and other African samples. Lastly, samples KS041, KDHTB008, KDHTB153 and KDHTB658 are classified as the rare N subtype which has previously been reported circulating within Africa [2, 6, 10]. Our reporting of four N subtypes within our samples from South Africa greatly adds to the number of sequences and sequence information currently available for this region. Other previously reported N subtype sequences include South African SAKS25 and SAKS30 as well as HKS35 from Uganda (Table 3). The general trend observed is that great variability is observed between subtypes, but within each subtype branch, variability is sparse. Within the respective subtypes, P, M and N, South African sequences lie in close proximity to each other. However, like with our South African sequences, those from other countries also appear to lie in multiple branches. For example, HKS35, HKS18, HKS58, HKS11 and HKS49 are all Ugandan samples but are dispersed between all three branches. Therefore, the K15 gene is less suited as a predictor of geographical region for subtyping compared to K1, but less intra-subtype variation of samples from specific geographical regions is occurring within this terminal region as compared to K1.

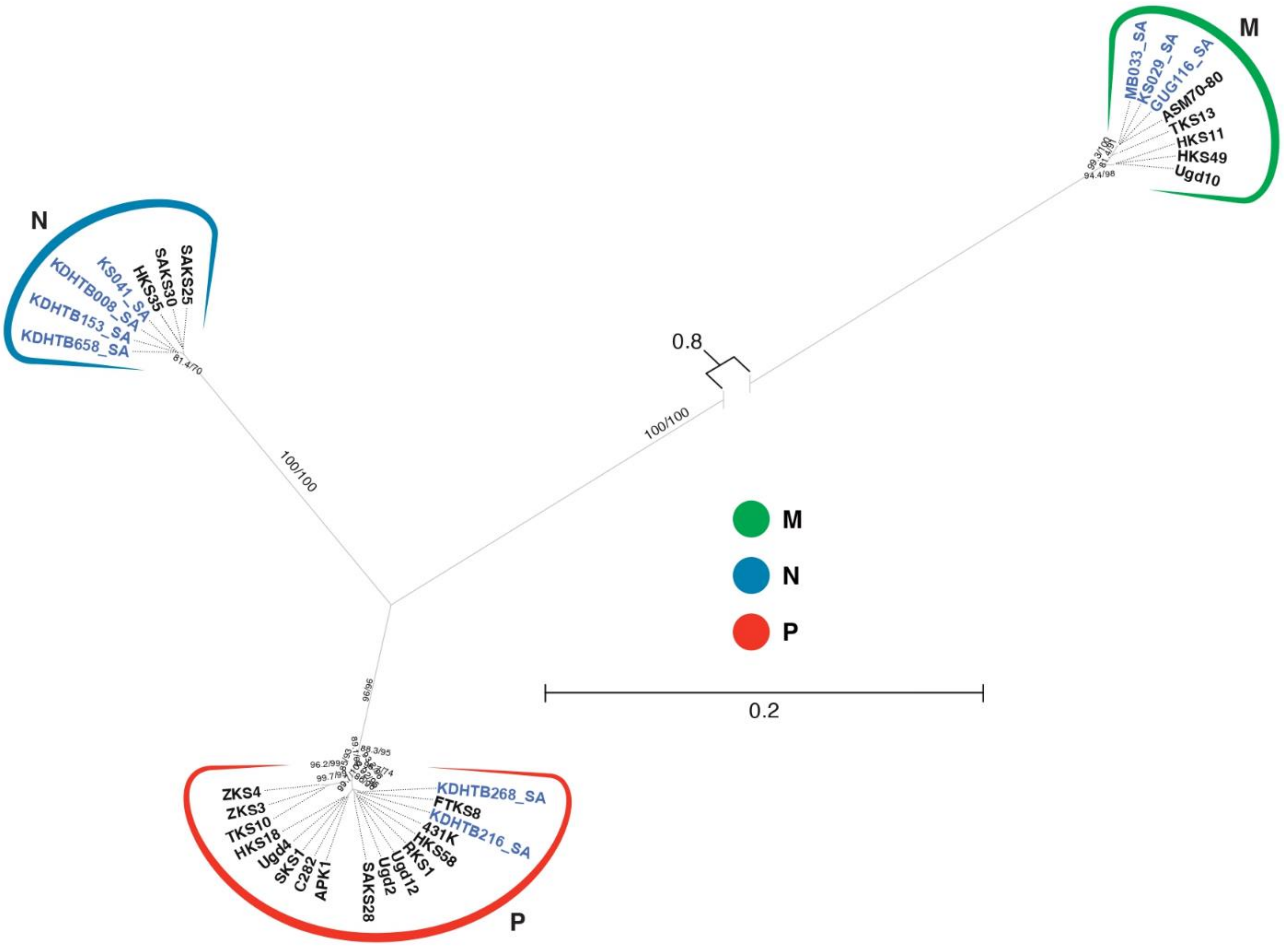


Figure 19: Maximum likelihood tree of the K15 gene region. The above phylogenetic tree illustrates sequence divergence and subtypes of the sequenced K15 South African sequences (indicated in blue and by '_SA') in relation to other subtypes from different geographical regions worldwide. This maximum likelihood tree was generated using selected reference sequences and iQtree software. Length of branches indicates sequence divergence between samples. Bootstrap values are provided on each branch for a 1000 bootstrap. Branches indicating K15 subtype are separated by colour, as per the key. Graphics courtesy of the Scientific Publications department (NIH/NCI).

A distance table was generated to investigate the percent identity between all nine K15 coding regions compared to the reference sample, GK18 (Table 5). We observed that both P subtypes, KDHTB216 and KDHTB268, display near 100% sequence similarity both in comparison to GK18 and between the two samples. This is due to all three being of the P subtype and would be expected. Samples KDHTB008, KDHTB153, KDHTB658 and KS041 are of the N subtype and display a lesser percent identity of approximately 70.4% on average compared to

GK18. In the situation of KDHTB008, a greater number of nucleotide differences occur as a result of a 1220 bp gap (55.8% of the coding region) in the K15 sequence likely as a result of poor viral load used for sequencing (see 4.2). Similarly, KS041 contains a 152 bp gap (7% of the coding region), likely as a result of a low viral load; however, the possibility of a true indel cannot be ruled out (Figure 20). Both observations occur in their respective reference guided alignments, which confirm it is not an error in the constructed de novo pipeline. As mentioned previously, there is a high degree of genetic divergence between subtypes, and subtypes can still be called on a partial sequence as unique subtype identifiers are still present. Lastly, the least percentage identity and hence lowest sequence similarity is shown with GUG116, KS029 and MB0333 which are all part of the most divergent M branch. An average of 45% percent identity to GK18 is observed here, thus illustrating the high degree of variance between these two subtypes which is consistent with previous observations [51, 211].

To conclude, the observed percent identities are to be expected and, in addition, confirm the above-mentioned K15 subtyping (Figure 19).

Table 5: Distance table depicting the percent identity (%) of the South African samples of this study compared to NC_009333.1 (GK18) for the K15 coding region. Shown is the percent identity, or similarity, of all South African sequences compared to GK18 as well as each other. KDHTB216 and KDHTB268 are P subtypes; KDHTB008, KDHTB153, KDHTB658 and KS041 are N subtypes; and GUG116, KS029 and MB033 are of the M subtype**.** The percentage identities are indicated as per the colour scheme.

GK18										*Percent Identity*
KDHTB216	98.75%									100.00%
KDHTB268	98.71%	99.86%								80.00%
KDHTB008	48.80%	48.96%	48.96%							60.00%
KDHTB153	78.74%	79.44%	79.35%	56.51%						40.00%
KDHTB658	78.78%	79.49%	79.40%	56.51%	99.95%					20.00%
KS041	75.44%	76.10%	76.00%	56.51%	94.58%	94.58%				0.00%
GUG116	45.40%	45.77%	45.77%	33.80%	44.87%	44.92%	43.41%			
KS029	45.36%	45.73%	45.73%	33.85%	44.92%	44.96%	43.45%	99.95%		
MB033	45.36%	45.73%	45.73%	33.85%	44.92%	44.96%	43.45%	99.95%	100.00%	
	GK18	**KDHTB216**	**KDHTB268**	**KDHTB008**	**KDHTB153**	**KDHTB658**	**KS041**	**GUG116**	**KS029**	**MB033**

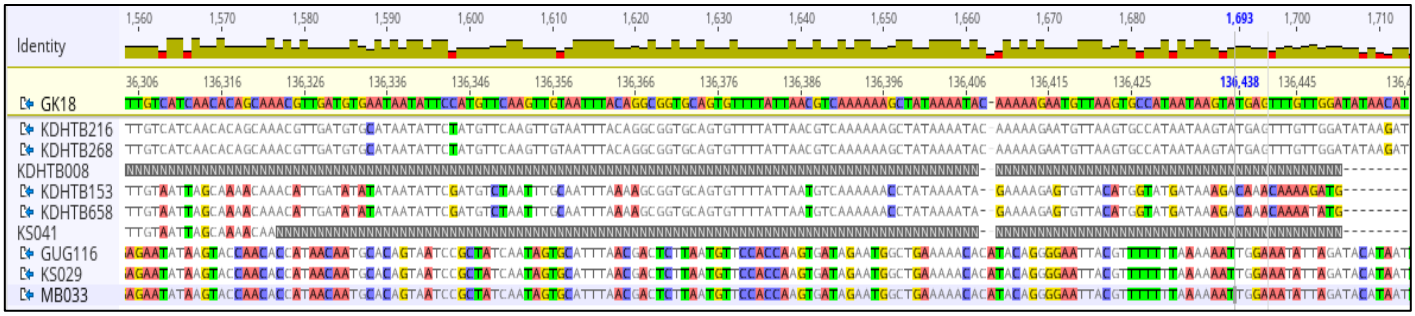


Figure 20: K15 alignment illustrating a portion of the gene whereby a 1220 bp gap is observed in KDHTB008 and a 152 bp gap for KS041. This may be due to poor coverage or a true indel, however in the case of KDHTB008 it is likely poor coverage. To note, samples KDHTB216 and KDHTB268 are P subtypes; KDHTB008, KDHTB153, KDHTB658 and KS041 are of the N subtype; and GUG116, KS029 and MB033 are of the M subtype.

4.3.3 Near full-length genome sequences

A phylogenetic tree was constructed from the nine South African sequences of this study to confirm both subtype as well as to make generalized comparisons of sequence similarity to those obtained by other studies and countries (Figure 21). Sequence coverage was gathered from the reference-guided alignment of each sample, and the corresponding data can be found in Table 6. Statistics related to the comparative reference-guided assembly pipeline can be found within the supplementary material. To note, '_SA' has been added to the name of each sample sequenced in this study to identify it against other sequences of similar naming convention.

Table 6: NGS sequence coverage, base pair length and K1/K15 subtypes for nine near full-length genome KSHV sequences derived from this study. Shown is the mean coverage achieved and final, trimmed length in base pairs for each sample alongside its classified sub-types. Mean coverage was ascertained from reference guided alignments to GK18.

#	Sample	Mean coverage	Length (bp)	K1/K15
1	KDHTB153_SA	198.7	151 392	A5/N
2	KDHTB658_SA	193.9	151 392	A5/N
3	KDHTB216_SA	190.7	151 392	B2/P
4	KDHTB268_SA	1786.4	151 392	B2/P
5	KDHTB008_SA	375.5	151 392	B2/N
6	KS029_SA	290.5	151 392	A5/M
7	KS041_SA	999.3	151 392	A5/N
8	MB033_SA	10391.8	151 392	A5/M
9	GUG116_SA	1683.0	151 392	A5/M

A number of whole-genome sequences were selected from other KSHV subtyping studies and mainly from the African continent making sure to capture the array of subtypes within each region (Table 3). This included well known Zambian sequences from Olp *et al* [6], Cameroonian samples from Marshall *et al* [2], Malawian samples from Moorad *et al* [205] as well as Ugandan sequences reported by Santiago *et al* [121] and Sallah *et al* [171]. For greater phylogenetic balance, some non-African samples were included from Japan [203], France [143], USA [121, 179, 212] and Greece [136].

The resulting phylogenetic tree (Figure 21) illustrates sequence diversity and recombination via splits and parallels drawn between branches with bootstrap values indicating how many times that branch was drawn in that manner. Importantly, K15 is the driving force behind the tree topology with K1 subtypes lying within each branch. This phylogenetic tree allows for a broader comparison of the KSHV genome as it depicts nearly every gene region within the KSHV genome apart from the repeat regions. In addition, a Phi test of recombination was performed within the SplitsTree software and found statistically significant evidence of recombination within the circulating South African KSHV strains ($p = 0.0$). Bootstrap values were included in tree generation and indicate a high degree of confidence (>80%) within branches. There is a high degree of diversity within the K15 subtypes as noted by the tree topology. Although there are more K1 subtypes reported in literature, these display subtype variability to a lesser degree than in comparison to K15 (as seen in 4.3.1 and 4.3.2 above) but are still distinct within each K15 branch as depicted in the figure. Geographical signal is clearly observed within each branch, whereby KSHV genome sequences from the same region generally tend to cluster together.

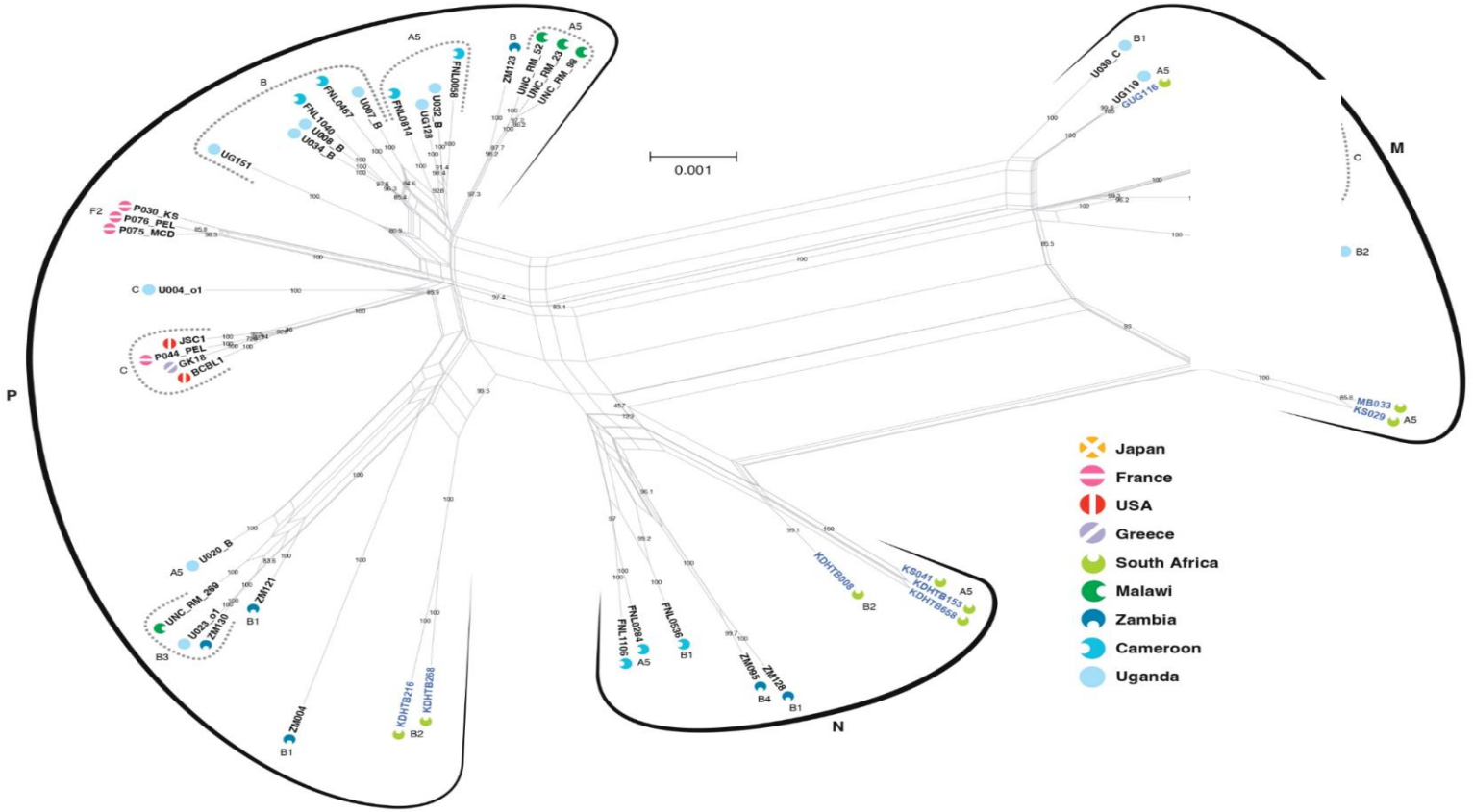


Figure 21: Unrooted nucleotide neighbour-net SplitsTree of all 9 obtained near full-length sequences from this study in comparison to reference sequences selected globally. The above phylogenetic tree illustrates sequence subtypes as well as the level of recombination occurring between all sequences included through the use of parallels drawn between sequences. Tree topology is defined by K15 subtype (M, N and P), and K1 subtype (A-F) is indicated within each branch. Length of branches indicates phylogenetic distances between samples and thus an indication of sequence divergence. Bootstrap values are provided on each branch for a 1000 bootstrap. South African samples sequenced in this study are indicated in blue and end in '_SA'. Countries have separated icons for ease of viewing. Several whole genome sequences were chosen to illustrate sequence diversity around the world. Graphics courtesy of the Scientific Publications department (NIH/NCI).

There are two South African sequences lying within the K15 P branch, KDHTB216_SA and KDHTB268_SA, which cluster together independently to the other references. Similarly, this phenomenon is observed in the N and M branches as well, confirming there is indeed a unique KSHV sequence variability occurring within this region. Strikingly, South African sequences clearly display a level of recombination between themselves, as sequences do not cluster as close together relative to other sequences, such as those from France and Japan. This is presumably due to the accumulation of variations within the K1 and central portion of the viral genome which distinguish them from sequences stemming from other regions. Alternatively, high seropositivity within the region may result in greater genetic variation.

Apart from studying K1 and K15 for subtyping purposes, it is important to make note of other significant observations within the KSHV genome. During quality review, it was noted that seven of the nine South African samples presented with one or more early stop codons. On further exploration, it was discovered that upstream frameshift mutations were the cause. KDHTB216_SA presented with a frameshift mutation within K4.2 whereby an additional adenosine is present, as well as in ORF48 and K8. Nonsynonymous mutations within K4.2 have previously been reported by Olp *et al* [6]. Furthermore, KDHTB153_SA presented with a missense mutation within the K7 gene in which a thymine base produced a stop codon. This observation is unique as early stop codons are not typically observed within this gene region. Lastly, six sequences contained frameshift mutations within the K8 gene resulting in downstream early stop codons. Mutations in the adjacent ORF K8.1 have previously been reported in sequences of African origin [6, 121]; however, it is unclear whether these two mutation sites are related. The effect of these mutations and the implications they pose on protein synthesis and downstream signalling are still to be defined. A summary of the positions of the forementioned mutations are shown in Table 7.

Table 7: Nonsynonymous mutations detected in seven of the near full-length sequences generated in this study. Shown are affected gene region, gene function, position and type of mutation observed. *May be a sequencing related homopolymer.

Sample	Position (5' – 3')	Type of mutation	Gene	Function
KDHTB216_SA	c.259+A	Frameshift	K4.2	Encodes for protein that interacts with pERP1. Interferes with immunoglobulin secretion. Contributor to lytic replication. [213]
KDHTB153_SA	c.286C>T	Missense	K7	Encodes a mitochondrial protein which aids in apoptosis inhibition. [214]
KDTB216_SA	c.192delC	Frameshift*	ORF48	Function yet to be determined.
KDHTB008_SA	c.547+C	Frameshift	K8	Encodes for a basic leucine zipper (bZip). [215] Important in viral reactivation and de novo infection. [216]
KDHTB153_SA	c.547+C	Frameshift		
KDHTB216_SA	c.599delC	Frameshift		
KDHTB658_SA	c.547+CC	Frameshift		
KS041_SA	c.547+C	Frameshift		
GUG116_SA	c.547+C	Frameshift		

5. Discussion

This study investigated KSHV sequences of clinical samples derived from South African individuals, thereby contributing novel knowledge to the existing KSHV database. Our results largely correspond to previously published literature examining KSHV sequences from the region of SSA.

This study has successfully produced 29 novel K1 sequences and a further nine near full-length sequences derived from patients in the Western Cape of South Africa. Sequence similarity to reference genomes in addition to phylogenetic analysis of the 29 K1 sequence subtypes found the majority (26/29, 90%) to be of the A5 subtype, and the minority (3/9, 10%) of the B2 subtype. This corresponds to previous studies examining the KSHV strains circulating within Africa, namely those originally produced by Olp *et al* [6] in 2015, Isaacs *et al* in 2016 [133], Santiago *et al* [121] in 2021 and most recently Marshall *et al* [2] in 2022.

The resulting A5 sequences produced in this study fell predominantly amongst KSHV strains circulating within Uganda and Cameroon, as depicted in the K1 phylogenetic tree (Figure 17). Ugandan samples included those from studies conducted by Sallah *et al* [171], Fouchard *et al* [207] and Santiago *et al* [121], as well as Cameroon samples reported by Lacoste *et al* [172] and Marshall *et al* [2]. This observation suggests that the majority of A5 subtypes circulating within South African are near identical strains to that found within the central region of Africa. This is further confirmed with two samples from the Central African Republic [172, 207] also lying within this branch, but slightly closer to the outlier KS062_SA. In addition, a sample originating from another South African study by Isaacs *et al* [133], KS42, also lies within the closely assembled net of A5 sequences from this study, thus confirming geographic similarity of South African circulating K1 subtypes. The same phenomenon is observed for KS41 (published by Isaacs *et al*) in the B branch. To note, Isaacs et al reported 42 A5 subtypes out of a total of 86 K1 sequences. The remainder consisted of B2 (16/86, 18.6%), B1 (10/86, 11.6%), B3 (9/86, 10.5%), A1 (4/86, 4.7%) and A4 (1/86, 1.1%) [133]. In the B branch, where our samples KDHTB216_SA, KDHTB008_SA and KDHTB268_SA are located, we observe sequences from Zambia (ZM130, ZM095, ZM121) and Uganda (U023-o1, UgD1, UgD2, UG118), as well as stand-alone samples from both Togo (K1-37/E44) and Senegal (K1-52/Ali) produced from an earlier study by Lacoste *et al* [172]. Here, the authors sequenced a single Togo sample of B2

subtype and two Senegal samples of B2 and C subtype. This suggests that the same subtypes observed in our data may extend into West Africa, however, more sequences from this region would need to be gathered and sequenced.

In regard to KSHV subtyping, East and West African strains are phylogenetically closer than those of European origin [187]. As the B strain appears to be unique to Africa, and phylogenetically older, the KSHV genome likely evolved within the African continent and extended to Europe where it diverged into groups A and C within the last 10 000 years [187]. This is illustrated in the K1 phylogenetic tree (Figure 17) whereby Central African Republic sequence K1-20/Gon and Ugandan U004-o1 sit amongst European C subtypes, such as Iap3 and GK18. It has been reported that KSHV was present in Europe prior to the AIDS pandemic, due to unique variants of the KSHV genome [187].

As a result of next-generation sequencing, nine K15 coding regions were extracted from the near full-length genome data generated in our study and subtyped accordingly. A phylogenetic tree of K15 was constructed for visualization (Figure 19). All three known K15 subtypes were accounted for, with two sequences being of the P subtype, 3 of the M subtype and 4 of the rarer N subtype. The P and M subtypes were the first to be reported, with the N subtype being introduced by Alagiozoglou *et al* [173] in 2000 from a South African patient cohort. Since then, other studies have also reported this subtype circulating within the region [10, 125, 173]. In addition, N subtypes have also been reported in other African countries such as Cameroon [2] and Zambia [6]. The K15 gene shows great variability between the three subtypes, however, marginal variation is observed within each subtype branch. This is not the case with K1 which has a number of different subtypes with variation carrying into each branch as well. Even though the N subtype has largely been reported from Africa, multiple K15 subtypes can occur per region, suggesting that perhaps K1 is a better indicator of geographical subtype.

Our study successfully produced nine near full-length KSHV sequences - the first near full-length sequences to come out of South Africa. Inspection of the acquired sequences revealed a combination of different K1/K15 subtypes as visualized in Figure 21 and Table 6. A Phi test further suggested statistically strong evidence of genetic recombination. Tree topology is driven by the K15 subtypes, followed by the K1 subtypes dispersing within each K15 branch. This is due to the amount of variation occurring in these regions as the central genome has

remained largely unchanged. In view of the fact that only a number of selected references were included, one can observe that South African samples appear to lie near those of Zambian origin within the full genome tree. This could be as a result of country proximity and genetic interchange between the herpesvirus in this region. This observation is seen in the P branch with samples KDHTB216_SA and KDHTB268_SA lying in close proximity to Zambian samples ZM004, ZM121 and ZM130. We can see that there is a single downstream branch separating these two samples, suggesting they are near identical in sequence identity. Looking towards the N branch, KS041_SA, KDHTB153_SA and KDHTB658_SA lie in closer proximity to each other than to KDHTB008_SA, suggesting sequence variance. However, this observation may be due to the poor coverage gap in K15 of sample KDHTB008_SA. Interestingly, we observe that two samples (KDHTB216_SA and KDTB268_SA) of the 3 reported B2 subtypes are found within the K15 B branch. The third B2 subtype (KDHTB008_SA) sits singularly in the K15 N subtype branch with the remaining three N subtype sequences, KS041_SA, KDHTB153_SA and KDHTB658_SA all being of the A5 subtype. Again, in the N branch of this tree, we can see that the Cameroon samples sit separately from the South African samples with the Zambian samples showing greater similarity. This phenomenon is also observed in the M branch where the two South African samples, MB033_SA and KS029_SA, are distanced from Ugandan samples taken from two separate studies (UG118 and UG119 [171]; U030_C [204]). The exception being one South African sample, GUG116_SA, which shows greater similarity to Ugandan sample UG119 than the other South African samples. This suggests that this individual may have been exposed to a Ugandan KSHV strain during the course of their life. This sample in particular also appears to be the most diverse of the whole-genome sequences from this study.

It can be observed that there is great sequence diversity between samples of African origin as opposed to those from France, Japan and the USA, which may be as a result of high viral recombination occurring within the African populations. As many African countries experience high degrees of poverty resulting in close living quarters and food sharing, high seroprevalence is expected. Not only this, but the African continent consists of many border-sharing countries, making human migration (such as for economic/political reasons) possible. Secondly, the high prevalence of many other viral pathogens, such as HIV and associated immunosuppression [217], may put adaptive pressure on this virus.

It is interesting to note that the A5 subtype is largely unique to SSA. Subtypes A1 to A4 are typically observed in European countries and differ substantially to the A5 sequence, which in turn rather displays distinct similarity to the SSA B subtype [125]. It is hypothesized that the A5 strain began to spread rapidly within SSA around the same time Bantu expansion took place from Western Africa approximately 4000 years ago [10].

When determining K1 and K15 subtypes, next-generation sequencing data is considered more reliable than that of Sanger sequencing data, due to numerous sequence contigs being produced per region with the average taken as the consensus. In addition, gene regions lying between these two terminals can be investigated further. Both the K1 and K15 gene regions are beneficial to obtain when sequencing a sample due to their hypervariability.

Subtype variability may be linked to KSHV-associated disease onset and progression. In the South African study by Isaacs *et al* [133], they observed that AIDS-KS patients harbouring the A5 subtype presented with more extensive disease, whilst those with the A1 and A4 subtypes showed a decreased likelihood of lower limb involvement, which is typical for classic KS. Another study by Mancuso *et al* [144] investigated classic KS within an Italian population. They reported that patients presenting with subtype A had faster disease progression than those of subtype C. Moreover, elevated viral loads were also observed in the A subtype. Additionally, another study examined KS lesions in a Brazilian population where subtypes A and C also dominated the KSHV pool. Interestingly, they observed that a significant portion of A subtypes were HIV positive whilst those with the C subtype were HIV negative, suggesting a non-random distribution in KS [218]. More extensive studies will be required to accurately determine a link between subtype and disease progression, especially in those with HIV/AIDS as these forementioned studies consisted of a small sample group. Our study did not follow up patients for disease outcome to observe any potential subtype correlations. Moreover, our study has also produced a limited sequence pool. We do, however, have a mix of individuals who presented with and without KS. All three B2 subtypes (KDHTB008_SA, KDHTB216_SA and KDHTB268_SA) where taken from the KDHTB cohort which did not have confirmed KS. Yet, no conclusions can suitably be drawn.

The major success of this study is the sequencing of nine near full-length genome sequences of South African origin. Near full-length sequences from Zambia, Cameroon, Uganda, Malawi and only a small quantity of others, but not from South Africa, have been produced from the

African continent, making this study an important addition to the KSHV research field. A limitation of this study is that no country of birth was able to be attained from the patients included. Although samples were taken from hospitals in South Africa, the individuals may have been born and raised elsewhere in Africa and acquired infection during childhood [155, 156]. Another limitation of this study was that our sample pool had previously been examined in other studies, therefore limiting the volume of concentrated viral DNA available for sequencing so that only 9 out of 57 samples qualified for NGS. For future studies, a greater sequencing output could be achieved using a richer source of KSHV, such as saliva, which has shown to contain higher viral loads [219, 220], and examining cohorts with a variety of KSHV-associated pathologies (such as KS, PEL, MCD and KICS) to potentially associate KSHV subtypes with disease outcome. Also, we found that the 3 μg enrichment protocol paired with the Illumina NextSeq sequencer as opposed to the previous generation MiSeq successfully sequenced samples with slightly lower concentration/viral load than required, which is an improvement of the methodology to be applied in future sequencing studies where limited sample quantities are available.

Lastly, it is important to note that not only were novel sequences derived from this study, but also training and invaluable knowledge was passed on from the United States to South Africa. With this technological transfer of skills and methods, we can broaden the landscape of viral sequencing in South Africa, providing a foundation for those who wish to venture into this aspect of research in the future.

6. Conclusion

In conclusion, this study successfully sequenced a total of 29 novel samples, thereby adding information on circulating KSHV K1 subtypes within the Western Cape Region of South Africa. Moreover, this study produced a total of nine near full-length KSHV sequences from this sample group. The A5 subtype proved to be the dominant subtype, making up 90% of the sequences, with the remainder being of the B2 subtype. Both subtypes have previously been reported from a separate South African study as well as in other KSHV studies derived from African individuals. In addition, K15 subtype could be called from the nine near full-length genome sequences produced which further adds to the limited information on KSHV K15 variety available for Southern Africa. Interestingly, all three K15 subtypes, P (22%), M (33%) and N (44%), were found to be in circulation. This confirms that the rarely reported N subtype indeed circulates within Africa, and in this case, the Southern region of the continent. Finally, phylogenetic analysis of the nine KSHV near full-length sequences revealed strong topology towards the K15 gene subtype, sequence similarity to other African sequences and sample variance within each branch. The produced phylogenetic tree allowed visualization of the complexity within this virus and the diversity of subtypes globally. Here we see where our samples sit in relation to others published as well as confirm subtype.

This study produced the first near full-length KSHV sequences originating from South Africa, which is a great achievement in the KSHV domain. Next-generation sequencing capabilities are relatively new in the field of science, and KSHV strains present worldwide are still actively being sequenced. With this knowledge, we hope that not only better understanding of the virus's migration patterns and evolutionary trends can be determined, but that the whole-genome sequences available may play a role in understanding KSHV-associated disease onset and progression, and perhaps treatment, in the future.

7. References

1. Broussard, G. and B. Damania, *Regulation of KSHV Latency and Lytic Reactivation.* Viruses, 2020. **12**(9).
2. Marshall, V.A., et al., *Systematic analysis of Kaposi's sarcoma (KS)-associated herpesvirus genomes from a KS case-control study in Cameroon: Evidence of dual infections but no association between viral sequence variation and KS risk.* Int J Cancer, 2022. **151**(7): p. 1127-1141.
3. Lopes, A.O., et al., *Update of the global distribution of human gammaherpesvirus 8 genotypes.* Sci Rep, 2021. **11**(1): p. 7640.
4. Mamimandjiami, A.I., et al., *Epidemiology and Genetic Variability of HHV-8/KSHV among Rural Populations and Kaposi's Sarcoma Patients in Gabon, Central Africa. Review of the Geographical Distribution of HHV-8 K1 Genotypes in Africa.* Viruses, 2021. **13**(2).
5. Lange, P. and B. Damania, *Kaposi Sarcoma-Associated Herpesvirus (KSHV).* Trends Microbiol, 2020. **28**(3): p. 236-237.
6. Olp, L.N., et al., *Whole-Genome Sequencing of Kaposi's Sarcoma-Associated Herpesvirus from Zambian Kaposi's Sarcoma Biopsy Specimens Reveals Unique Viral Diversity.* Journal of Virology, 2015. **89**(24): p. 12299-12308.
7. Majerciak, V., et al., *A Viral Genome Landscape of RNA Polyadenylation from KSHV Latent to Lytic Infection.* PLoS pathogens, 2013. **9**: p. e1003749.
8. Jones, D., et al., *Primary-Effusion Lymphoma and Kaposi's Sarcoma in a Cardiac-Transplant Recipient.* New England Journal of Medicine, 1998. **339**(7): p. 444-449.
9. Mehta, S., et al., *Kaposi's sarcoma as a presenting manifestation of HIV.* Indian J Sex Transm Dis AIDS, 2011. **32**(2): p. 108-10.
10. Hayward, G.S. and J.C. Zong, *Modern evolutionary history of the human KSHV genome.* Curr Top Microbiol Immunol, 2007. **312**: p. 1-42.
11. Zhou, T., et al., *Multicentric Castleman disease and the evolution of the concept.* Pathologica, 2021. **113**(5): p. 339-353.
12. Jary, A., et al., *Kaposi's Sarcoma-Associated Herpesvirus, the Etiological Agent of All Epidemiological Forms of Kaposi's Sarcoma.* Cancers (Basel), 2021. **13**(24).
13. Chang, Y., et al., *Identification of Herpesvirus-Like DNA Sequences in AIDS-Associated Kaposi's Sarcoma.* Science, 1994. **266**(5192).
14. Dittmer, D.P. and B. Damania, *Kaposi's Sarcoma-Associated Herpesvirus (KSHV)-Associated Disease in the AIDS Patient: An Update.* Cancer Treat Res, 2019. **177**: p. 63-80.
15. Motlhale, M., et al., *Epidemiology of Kaposi's sarcoma in sub-Saharan Africa.* Cancer Epidemiol, 2022. **78**: p. 102167.
16. Martin, J.N., et al., *Sexual transmission and the natural history of human herpesvirus 8 infection.* N Engl J Med, 1998. **338**(14): p. 948-54.
17. Gessain, A., et al., *Human herpesvirus 8 primary infection occurs during childhood in Cameroon, Central Africa.* Int J Cancer, 1999. **81**(2): p. 189-92.
18. Malope-Kgokong, B.I., et al., *Kaposi's Sarcoma Associated-Herpes Virus (KSHV) Seroprevalence in Pregnant Women in South Africa.* Infect Agent Cancer, 2010. **5**: p. 14.
19. Mbulaiteye, S., et al., *Molecular evidence for mother-to-child transmission of Kaposi sarcoma-associated herpesvirus in Uganda and K1 gene evolution within the host.* J Infect Dis, 2006. **193**(9): p. 1250-7.

20. Lebbe, C., et al., *Diagnosis and treatment of Kaposi's sarcoma: European consensus-based interdisciplinary guideline (EDF/EADO/EORTC).* Eur J Cancer, 2019. **114**: p. 117-127.
21. Addula, D., C.J. Das, and V. Kundra, *Imaging of Kaposi sarcoma.* Abdom Radiol (NY), 2021. **46**(11): p. 5297-5306.
22. Cancer.net. *Sarcoma -Kaposi: Statistics*. Sarcoma 2023 03/2023 [cited 2023 05/10]; KS Statistics]. Available from: https://www.cancer.net/cancer-types/sarcoma-kaposi/statistics.
23. WHO. *Kaposi Sarcoma Fact Sheet*. 2020 [cited 2023 19/05]; Available from: https://gco.iarc.fr/today/data/factsheets/cancers/19-Kaposi-sarcoma-fact-sheet.pdf.
24. Cesarman, E., et al., *In vitro establishment and characterization of two acquired immunodeficiency syndrome-related lymphoma cell lines (BC-1 and BC-2) containing Kaposi's sarcoma-associated herpesvirus-like (KSHV) DNA sequences.* Blood, 1995. **86**(7): p. 2708-14.
25. Soulier, J., et al., *Kaposi's sarcoma-associated herpesvirus-like DNA sequences in multicentric Castleman's disease.* Blood, 1995. **86**(4): p. 1276-80.
26. Polizzotto, M.N., et al., *Clinical Features and Outcomes of Patients With Symptomatic Kaposi Sarcoma Herpesvirus (KSHV)-associated Inflammation: Prospective Characterization of KSHV Inflammatory Cytokine Syndrome (KICS).* Clin Infect Dis, 2016. **62**(6): p. 730-738.
27. Hernández-Ramírez, R.U., et al., *Cancer risk in HIV-infected people in the USA from 1996 to 2012: a population-based, registry-linkage study.* Lancet HIV, 2017. **4**(11): p. e495-e504.
28. Hleyhel, M., et al., *Risk of AIDS-defining cancers among HIV-1-infected patients in France between 1992 and 2009: results from the FHDH-ANRS CO4 cohort.* Clin Infect Dis, 2013. **57**(11): p. 1638-47.
29. Russo, J.J., et al., *Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8).* Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14862-7.
30. Melendez, L.V., et al., *An apparently new herpesvirus from primary kidney cultures of the squirrel monkey (Saimiri sciureus).* Lab Anim Care, 1968. **18**(3): p. 374-81.
31. Melendez, L.V., et al., *Herpesvirus ateles, a new lymphoma virus of monkeys.* Nat New Biol, 1972. **235**(58): p. 182-4.
32. Desrosiers, R.C., et al., *A herpesvirus of rhesus monkeys related to the human Kaposi's sarcoma-associated herpesvirus.* J Virol, 1997. **71**(12): p. 9764-9.
33. Schultz, E.R., et al., *Characterization of two divergent lineages of macaque rhadinoviruses related to Kaposi's sarcoma-associated herpesvirus.* J Virol, 2000. **74**(10): p. 4919-28.
34. Bosch, M.L., K.B. Strand, and T.M. Rose, *Gammaherpesvirus sequence comparisons.* J Virol, 1998. **72**(10): p. 8458-9.
35. Alexander, L., et al., *The primary sequence of rhesus monkey rhadinovirus isolate 26-95: sequence similarities to Kaposi's sarcoma-associated herpesvirus and rhesus monkey rhadinovirus isolate 17577.* J Virol, 2000. **74**(7): p. 3388-98.
36. Greensill, J., et al., *Two distinct gamma-2 herpesviruses in African green monkeys: a second gamma-2 herpesvirus lineage among old world primates?* J Virol, 2000. **74**(3): p. 1572-7.
37. Whitley, R.J., *Herpesviruses*, in *Medical Microbiology*, S. Baron, Editor. 1996: Galveston (TX).

38. Nealon, K., et al., *Lytic replication of Kaposi's sarcoma-associated herpesvirus results in the formation of multiple capsid species: isolation and molecular characterization of A, B, and C capsids from a gammaherpesvirus.* J Virol, 2001. **75**(6): p. 2866-78.
39. Zhu, F.X., et al., *Virion proteins of Kaposi's sarcoma-associated herpesvirus.* J Virol, 2005. **79**(2): p. 800-11.
40. Yan, L., et al., *Towards Better Understanding of KSHV Life Cycle: from Transcription and Posttranscriptional Regulations to Pathogenesis.* Virol Sin, 2019. **34**(2): p. 135-161.
41. Sathish, N., X. Wang, and Y. Yuan, *Tegument Proteins of Kaposi's Sarcoma-Associated Herpesvirus and Related Gamma-Herpesviruses.* Front Microbiol, 2012. **3**: p. 98.
42. Akula, S.M., et al., *Integrin alpha3beta1 (CD 49c/29) is a cellular receptor for Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) entry into the target cells.* Cell, 2002. **108**(3): p. 407-19.
43. Spear, P.G. and R. Longnecker, *Herpesvirus entry: an update.* J Virol, 2003. **77**(19): p. 10179-85.
44. Birkmann, A., et al., *Cell surface heparan sulfate is a receptor for human herpesvirus 8 and interacts with envelope glycoprotein K8.1.* J Virol, 2001. **75**(23): p. 11583-93.
45. Muniraju, M., et al., *Kaposi Sarcoma-Associated Herpesvirus Glycoprotein H Is Indispensable for Infection of Epithelial, Endothelial, and Fibroblast Cell Types.* J Virol, 2019. **93**(16).
46. Rezaee, S.A.R., et al., *Kaposi's sarcoma-associated herpesvirus immune modulation: an overview.* J Gen Virol, 2006. **87**(Pt 7): p. 1781-1804.
47. Chen, W., et al., *K15 Protein of Kaposi's Sarcoma Herpesviruses Increases Endothelial Cell Proliferation and Migration through Store-Operated Calcium Entry.* Viruses, 2018. **10**(6).
48. Abere, B., et al., *The Kaposi's sarcoma-associated herpesvirus (KSHV) non-structural membrane protein K15 is required for viral lytic replication and may represent a therapeutic target.* PLoS Pathog, 2017. **13**(9): p. e1006639.
49. Wen, K.W. and B. Damania, *Kaposi sarcoma-associated herpesvirus (KSHV): molecular biology and oncogenesis.* Cancer Lett, 2010. **289**(2): p. 140-50.
50. Zong, J.C., et al., *High-level variability in the ORF-K1 membrane protein gene at the left end of the Kaposi's sarcoma-associated herpesvirus genome defines four major virus subtypes and multiple variants or clades in different human populations.* J Virol, 1999. **73**(5): p. 4156-70.
51. Poole, L.J., et al., *Comparison of genetic variability at multiple loci across the genomes of the major subtypes of Kaposi's sarcoma-associated herpesvirus reveals evidence for recombination and for two distinct types of open reading frame K15 alleles at the right-hand end.* J Virol, 1999. **73**(8): p. 6646-60.
52. Zong, J.C., et al., *Evaluation of global clustering patterns and strain variation over an extended ORF26 gene locus from Kaposi's sarcoma herpesvirus.* J Clin Virol, 2007. **40**(1): p. 19-25.
53. Dollery, S.J., *Towards Understanding KSHV Fusion and Entry.* Viruses, 2019. **11**(11).
54. Akula, S.M., et al., *Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) infection of human fibroblast cells occurs through endocytosis.* J Virol, 2003. **77**(14): p. 7978-90.
55. Inoue, N., et al., *Characterization of entry mechanisms of human herpesvirus 8 by using an Rta-dependent reporter cell line.* J Virol, 2003. **77**(14): p. 8147-52.

56. Kerur, N., et al., *Characterization of entry and infection of monocytic THP-1 cells by Kaposi's sarcoma associated herpesvirus (KSHV): role of heparan sulfate, DC-SIGN, integrins and signaling.* Virology, 2010. **406**(1): p. 103-16.
57. Kumar, B. and B. Chandran, *KSHV Entry and Trafficking in Target Cells-Hijacking of Cell Signal Pathways, Actin and Membrane Dynamics.* Viruses, 2016. **8**(11).
58. Campbell, M., et al., *Epigenetic Regulation of Kaposi's Sarcoma-Associated Herpesvirus Latency.* Front Microbiol, 2020. **11**: p. 850.
59. Xu, Y., et al., *A Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 ORF50 deletion mutant is defective for reactivation of latent virus and DNA replication.* J Virol, 2005. **79**(6): p. 3479-87.
60. Greene, W., et al., *Molecular biology of KSHV in relation to AIDS-associated oncogenesis.* Cancer Treat Res, 2007. **133**: p. 69-127.
61. Sun, R., et al., *A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus.* Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10866-71.
62. Guito, J. and D.M. Lukac, *KSHV Rta Promoter Specification and Viral Reactivation.* Front Microbiol, 2012. **3**: p. 30.
63. Ye, F., et al., *Reactive oxygen species hydrogen peroxide mediates Kaposi's sarcoma-associated herpesvirus reactivation from latency.* PLoS Pathog, 2011. **7**(5): p. e1002054.
64. Davis, D.A., et al., *Hypoxia induces lytic replication of Kaposi sarcoma-associated herpesvirus.* Blood, 2001. **97**(10): p. 3244-50.
65. Vieira, J., et al., *Activation of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) lytic replication by human cytomegalovirus.* J Virol, 2001. **75**(3): p. 1378-86.
66. Tang, Q., et al., *Herpes simplex virus type 2 triggers reactivation of Kaposi's sarcoma-associated herpesvirus from latency and collaborates with HIV-1 Tat.* PLoS One, 2012. **7**(2): p. e31652.
67. Jiang, Y., et al., *Mutual inhibition between Kaposi's sarcoma-associated herpesvirus and Epstein-Barr virus lytic replication initiators in dually-infected primary effusion lymphoma.* PLoS One, 2008. **3**(2): p. e1569.
68. Zhang, Z., et al., *The K1 Protein of Kaposi's Sarcoma-Associated Herpesvirus Augments Viral Lytic Replication.* J Virol, 2016. **90**(17): p. 7657-66.
69. Wang, L., et al., *Immortalization of primary endothelial cells by the K1 protein of Kaposi's sarcoma-associated herpesvirus.* Cancer Res, 2006. **66**(7): p. 3658-66.
70. Chandriani, S. and D. Ganem, *Array-based transcript profiling and limiting-dilution reverse transcription-PCR analysis identify additional latent genes in Kaposi's sarcoma-associated herpesvirus.* J Virol, 2010. **84**(11): p. 5565-73.
71. Lagunoff, M., et al., *Deregulated signal transduction by the K1 gene product of Kaposi's sarcoma-associated herpesvirus.* Proc Natl Acad Sci U S A, 1999. **96**(10): p. 5704-9.
72. Lee, B.S., et al., *Characterization of the Kaposi's sarcoma-associated herpesvirus K1 signalosome.* J Virol, 2005. **79**(19): p. 12173-84.
73. Wang, L., et al., *The Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) K1 protein induces expression of angiogenic and invasion factors.* Cancer Res, 2004. **64**(8): p. 2774-81.
74. Bala, K., et al., *Kaposi's sarcoma herpesvirus K15 protein contributes to virus-induced angiogenesis by recruiting PLCgamma1 and activating NFAT1-dependent RCAN1 expression.* PLoS Pathog, 2012. **8**(9): p. e1002927.

75. Katano, H., et al., *Expression and localization of human herpesvirus 8-encoded proteins in primary effusion lymphoma, Kaposi's sarcoma, and multicentric Castleman's disease.* Virology, 2000. **269**(2): p. 335-44.
76. Grundhoff, A. and D. Ganem, *Inefficient establishment of KSHV latency suggests an additional role for continued lytic replication in Kaposi sarcoma pathogenesis.* J Clin Invest, 2004. **113**(1): p. 124-36.
77. Parravicini, C., et al., *Differential viral protein expression in Kaposi's sarcoma-associated herpesvirus-infected diseases: Kaposi's sarcoma, primary effusion lymphoma, and multicentric Castleman's disease.* Am J Pathol, 2000. **156**(3): p. 743-9.
78. Vinh, D., A. Ahn, and N. Mohyuddin, *HIV-Positive Patient With an Obstructing Laryngeal Mass.* JAMA Otolaryngol Head Neck Surg, 2020. **146**(7): p. 667-668.
79. Barron, K., et al., *Kaposi Sarcoma of the Larynx: A Systematic Review.* Otolaryngol Head Neck Surg, 2023. **168**(3): p. 269-281.
80. Watson, J.R., D. Granoff, and R.T. Sataloff, *Dysphonia due to Kaposi's sarcoma as the presenting symptom of human immunodeficiency virus.* J Voice, 2004. **18**(3): p. 398-402.
81. Angouridakis, N., et al., *Classic (Mediterranean) Kaposi's sarcoma of the true vocal cord: a case report and review of the literature.* Eur Arch Otorhinolaryngol, 2006. **263**(6): p. 537-40.
82. Kim, T.H., et al., *Misdiagnosis of Human Herpes Virus-8-Associated Kaposi's Sarcoma as Adverse Drug Eruptions.* Arch Plast Surg, 2022. **49**(3): p. 457-461.
83. Aghazadeh, N., et al., *Kaposi sarcoma misdiagnosed as granuloma annulare: A case of mistaken identity.* J Cutan Pathol, 2021. **48**(2): p. 318-321.
84. Torrence, G.M. and J.S. Wrobel, *A case of mistaken identity: classic Kaposi sarcoma misdiagnosed as a diabetic foot ulcer in an atypical patient.* Clin Diabetes Endocrinol, 2019. **5**: p. 8.
85. Blumenthal, M.J., et al., *The Contribution of Kaposi's Sarcoma-Associated Herpesvirus to Mortality in Hospitalized Human Immunodeficiency Virus-Infected Patients Being Investigated for Tuberculosis in South Africa.* J Infect Dis, 2019. **220**(5): p. 841-851.
86. McCloskey, D., et al., *LAMP-enabled diagnosis of Kaposi's sarcoma for sub-Saharan Africa.* Sci Adv, 2023. **9**(2): p. eadc8913.
87. Schneider, J.W. and D.P. Dittmer, *Diagnosis and Treatment of Kaposi Sarcoma.* Am J Clin Dermatol, 2017. **18**(4): p. 529-539.
88. Motlhale, M., et al., *Kaposi sarcoma-associated herpesvirus, HIV-1 and Kaposi sarcoma risk in black South Africans diagnosed with cancer during antiretroviral treatment rollout.* Int J Cancer, 2023. **152**(10): p. 2081-2089.
89. Oluoch, P.O., et al., *Distinctive Kaposi Sarcoma-Associated Herpesvirus Serological Profile during Acute Plasmodium falciparum Malaria Episodes.* Int J Mol Sci, 2023. **24**(7).
90. Blumenthal, M.J., et al., *Evidence for altered host genetic factors in KSHV infection and KSHV-related disease development.* Rev Med Virol, 2021. **31**(2): p. e2160.
91. UNAIDS. *Prevalence of HIV, total (% of population ages 15-49) - Sub-Saharan Africa*. 2021 [cited 2023 20/5]; Available from: https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=ZG&view=map&year=2021.
92. Zuma, K., et al., *The HIV Epidemic in South Africa: Key Findings from 2017 National Population-Based Survey.* Int J Environ Res Public Health, 2022. **19**(13).

93. Government, S.A., *Mid-year population estimates 2021*. 2022, Statistics of South Africa: www.statssa.gov.za/publications/P0302/P03022021.pdf.
94. Oksenhendler, E., et al., *Multicentric Castleman's disease in HIV infection: a clinical and pathological study of 20 patients.* AIDS, 1996. **10**(1): p. 61-7.
95. Oksenhendler, E., et al., *High levels of human herpesvirus 8 viral load, human interleukin-6, interleukin-10, and C reactive protein correlate with exacerbation of multicentric castleman disease in HIV-infected patients.* Blood, 2000. **96**(6): p. 2069-73.
96. Polizzotto, M.N., et al., *Human and viral interleukin-6 and other cytokines in Kaposi sarcoma herpesvirus-associated multicentric Castleman disease.* Blood, 2013. **122**(26): p. 4189-98.
97. Bower, M., et al., *Cytokine changes during rituximab therapy in HIV-associated multicentric Castleman disease.* Blood, 2009. **113**(19): p. 4521-4.
98. Chadburn, A., et al., *HHV8/KSHV-Positive Lymphoproliferative Disorders and the Spectrum of Plasmablastic and Plasma Cell Neoplasms: 2015 SH/EAHP Workshop Report—Part 3.* American Journal of Clinical Pathology, 2017. **147**(2): p. 171-187.
99. Robey, R.C., et al., *The use of monoclonal antibodies to treat Castleman's disease.* Immunotherapy, 2014. **6**(2): p. 211-9.
100. Lurain, K., R. Yarchoan, and T.S. Uldrick, *Treatment of Kaposi Sarcoma Herpesvirus-Associated Multicentric Castleman Disease.* Hematol Oncol Clin North Am, 2018. **32**(1): p. 75-88.
101. Ramaswami, R., et al., *Characteristics and outcomes of KSHV-associated multicentric Castleman disease with or without other KSHV diseases.* Blood Adv, 2021. **5**(6): p. 1660-1670.
102. Schulz, T.F., *The pleiotropic effects of Kaposi's sarcoma herpesvirus.* J Pathol, 2006. **208**(2): p. 187-98.
103. Kaplan, L.D., *Human herpesvirus-8: Kaposi sarcoma, multicentric Castleman disease, and primary effusion lymphoma.* Hematology Am Soc Hematol Educ Program, 2013. **2013**: p. 103-8.
104. Knowles, D., et al., *Molecular genetic analysis of three AIDS-associated neoplasms of uncertain lineage demonstrates their B-cell derivation and the possible pathogenetic role of the Epstein-Barr virus.* Blood, 1989. **73**(3): p. 792-799.
105. Horenstein, M.G., et al., *Epstein-Barr Virus Latent Gene Expression in Primary Effusion Lymphomas Containing Kaposi's Sarcoma–Associated Herpesvirus/Human Herpesvirus-8.* Blood, 1997. **90**(3): p. 1186-1191.
106. Friborg, J., Jr., et al., *p53 inhibition by the LANA protein of KSHV protects against cell death.* Nature, 1999. **402**(6764): p. 889-94.
107. Radkov, S.A., P. Kellam, and C. Boshoff, *The latent nuclear antigen of Kaposi sarcoma-associated herpesvirus targets the retinoblastoma-E2F pathway and with the oncogene Hras transforms primary rat cells.* Nat Med, 2000. **6**(10): p. 1121-7.
108. Li, M., et al., *Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin.* J Virol, 1997. **71**(3): p. 1984-91.
109. Godden-Kent, D., et al., *The cyclin encoded by Kaposi's sarcoma-associated herpesvirus stimulates cdk6 to phosphorylate the retinoblastoma protein and histone H1.* J Virol, 1997. **71**(6): p. 4193-8.
110. Ellis, M., et al., *Degradation of p27(Kip) cdk inhibitor triggered by Kaposi's sarcoma virus cyclin-cdk6 complex.* EMBO J, 1999. **18**(3): p. 644-53.

111. Belanger, C., et al., *Human herpesvirus 8 viral FLICE-inhibitory protein inhibits Fas-mediated apoptosis through binding and prevention of procaspase-8 maturation.* J Hum Virol, 2001. **4**(2): p. 62-73.
112. Field, N., et al., *KSHV vFLIP binds to IKK-gamma to activate IKK.* J Cell Sci, 2003. **116**(Pt 18): p. 3721-8.
113. Liu, L., et al., *The human herpes virus 8-encoded viral FLICE inhibitory protein physically associates with and persistently activates the Ikappa B kinase complex.* J Biol Chem, 2002. **277**(16): p. 13745-51.
114. Fernández-Trujillo, L., et al., *Primary effusion lymphoma in a human immunodeficiency virus-negative patient with unexpected unusual complications: a case report.* Journal of Medical Case Reports, 2019. **13**(1): p. 301.
115. Uldrick, T.S., M.N. Polizzotto, and R. Yarchoan, *Recent advances in Kaposi sarcoma herpesvirus-associated multicentric Castleman disease.* Curr Opin Oncol, 2012. **24**(5): p. 495-505.
116. Uldrick, T.S., et al., *An interleukin-6-related systemic inflammatory syndrome in patients co-infected with Kaposi sarcoma-associated herpesvirus and HIV but without Multicentric Castleman disease.* Clin Infect Dis, 2010. **51**(3): p. 350-8.
117. Pineiro, F., et al., *Clinical and pathological features of Kaposi sarcoma herpesvirus-associated inflammatory cytokine syndrome.* AIDS, 2020. **34**(14): p. 2097-2101.
118. Cantos, V.D., et al., *Experience with Kaposi Sarcoma Herpesvirus Inflammatory Cytokine Syndrome in a Large Urban HIV Clinic in the United States: Case Series and Literature Review.* Open Forum Infect Dis, 2017. **4**(4): p. ofx196.
119. Polizzotto, M.N., et al., *18F-fluorodeoxyglucose Positron Emission Tomography in Kaposi Sarcoma Herpesvirus-Associated Multicentric Castleman Disease: Correlation With Activity, Severity, Inflammatory and Virologic Parameters.* J Infect Dis, 2015. **212**(8): p. 1250-60.
120. Polizzotto, M.N., et al., *Clinical Manifestations of Kaposi Sarcoma Herpesvirus Lytic Activation: Multicentric Castleman Disease (KSHV-MCD) and the KSHV Inflammatory Cytokine Syndrome.* Front Microbiol, 2012. **3**: p. 73.
121. Santiago, J.C., et al., *Intra-host changes in Kaposi sarcoma-associated herpesvirus genomes in Ugandan adults with Kaposi sarcoma.* PLoS Pathog, 2021. **17**(1): p. e1008594.
122. Gunther, T., et al., *Influence of ND10 components on epigenetic determinants of early KSHV latency establishment.* PLoS Pathog, 2014. **10**(7): p. e1004274.
123. Gunther, T. and A. Grundhoff, *The epigenetic landscape of latent Kaposi sarcoma-associated herpesvirus genomes.* PLoS Pathog, 2010. **6**(6): p. e1000935.
124. Tomlinson, C.C. and B. Damania, *The K1 protein of Kaposi's sarcoma-associated herpesvirus activates the Akt signaling pathway.* J Virol, 2004. **78**(4): p. 1918-27.
125. Zong, J., et al., *Genotypic analysis at multiple loci across Kaposi's sarcoma herpesvirus (KSHV) DNA molecules: clustering patterns, novel variants and chimerism.* J Clin Virol, 2002. **23**(3): p. 119-48.
126. McGeoch, D.J., *Molecular evolution of the gamma-Herpesvirinae.* Philos Trans R Soc Lond B Biol Sci, 2001. **356**(1408): p. 421-35.
127. Biggar, R.J., et al., *Human herpesvirus 8 in Brazilian Amerindians: a hyperendemic population with a new subtype.* J Infect Dis, 2000. **181**(5): p. 1562-8.
128. Stebbing, J., et al., *Significance of variation within HIV, EBV, and KSHV subtypes.* J Int Assoc Physicians AIDS Care (Chic), 2006. **5**(3): p. 93-102.

129. An, W. and A. Telesnitsky, *HIV-1 genetic recombination: experimental approaches and observations.* AIDS Rev, 2002. **4**(4): p. 195-212.
130. Nicholas, J., et al., *Novel organizational features, captured cellular genes, and strain variability within the genome of KSHV/HHV8.* J Natl Cancer Inst Monogr, 1998(23): p. 79-88.
131. Tornesello, M.L., et al., *Human herpesvirus type 8 variants circulating in Europe, Africa and North America in classic, endemic and epidemic Kaposi's sarcoma lesions during pre-AIDS and AIDS era.* Virology, 2010. **398**(2): p. 280-9.
132. Kajumbula, H., et al., *Ugandan Kaposi's sarcoma-associated herpesvirus phylogeny: evidence for cross-ethnic transmission of viral subtypes.* Intervirology, 2006. **49**(3): p. 133-43.
133. Isaacs, T., et al., *Genetic diversity of HHV8 subtypes in South Africa: A5 subtype is associated with extensive disease in AIDS-KS.* J Med Virol, 2016. **88**(2): p. 292-303.
134. Hayward, G.S. *KSHV strains: the origins and global spread of the virus*. in *Seminars in cancer biology*. 1999. Elsevier.
135. Gramolelli, S. and T.F. Schulz, *The role of Kaposi sarcoma-associated herpesvirus in the pathogenesis of Kaposi sarcoma.* J Pathol, 2015. **235**(2): p. 368-80.
136. Glenn, M., et al., *Identification of a spliced gene from Kaposi's sarcoma-associated herpesvirus encoding a protein with similarities to latent membrane proteins 1 and 2A of Epstein-Barr virus.* J Virol, 1999. **73**(8): p. 6953-63.
137. Brinkmann, M.M., et al., *Activation of mitogen-activated protein kinase and NF-kappaB pathways by a Kaposi's sarcoma-associated herpesvirus K15 membrane protein.* J Virol, 2003. **77**(17): p. 9346-58.
138. Choi, J.K., et al., *Identification of the novel K15 gene at the rightmost end of the Kaposi's sarcoma-associated herpesvirus genome.* J Virol, 2000. **74**(1): p. 436-46.
139. Havemeier, A., et al., *Activation of NF-kappaB by the Kaposi's sarcoma-associated herpesvirus K15 protein involves recruitment of the NF-kappaB-inducing kinase, IkappaB kinases, and phosphorylation of p65.* J Virol, 2014. **88**(22): p. 13161-72.
140. Stebbing, J., et al., *Lack of intra-patient strain variability during infection with Kaposi's sarcoma-associated herpesvirus.* Am J Hematol, 2001. **68**(2): p. 133-4.
141. Duprez, R., et al., *Cutaneous disseminated endemic Kaposi's sarcoma in a Polynesian man infected with a new divergent human herpesvirus 8 subtype D.* J Clin Virol, 2006. **37**(3): p. 222-6.
142. Betsem, E., et al., *Epidemiology and genetic variability of HHV-8/KSHV in Pygmy and Bantu populations in Cameroon.* PLoS Negl Trop Dis, 2014. **8**(5): p. e2851.
143. Jary, A., et al., *New Kaposi's sarcoma-associated herpesvirus variant in men who have sex with men associated with severe pathologies.* J Infect Dis, 2020. **222**(8): p. 1320-1328.
144. Mancuso, R., et al., *HHV8 a subtype is associated with rapidly evolving classic Kaposi's sarcoma.* J Med Virol, 2008. **80**(12): p. 2153-60.
145. Inoue, N., et al., *New immunofluorescence assays for detection of Human herpesvirus 8-specific antibodies.* Clin Diagn Lab Immunol, 2000. **7**(3): p. 427-35.
146. Lam, L.L., et al., *Highly sensitive assay for human herpesvirus 8 antibodies that uses a multiple antigenic peptide derived from open reading frame K8.1.* J Clin Microbiol, 2002. **40**(2): p. 325-9.

147. Mbisa, G.L., et al., *Detection of antibodies to Kaposi's sarcoma-associated herpesvirus: a new approach using K8.1 ELISA and a newly developed recombinant LANA ELISA.* J Immunol Methods, 2010. **356**(1-2): p. 39-46.
148. Watanabe, D., et al., *Evaluation of human herpesvirus-8 viremia and antibody positivity in patients with HIV infection with human herpesvirus-8-related diseases.* Journal of Medical Virology, 2023. **95**(12): p. e29324.
149. Labo, N., et al., *Heterogeneity and Breadth of Host Antibody Response to KSHV Infection Demonstrated by Systematic Analysis of the KSHV Proteome.* PLOS Pathogens, 2014. **10**(3): p. e1004046.
150. Sabourin, K.R., et al., *Malaria Is Associated With Kaposi Sarcoma-Associated Herpesvirus Seroconversion in a Cohort of Western Kenyan Children.* J Infect Dis, 2021. **224**(2): p. 303-311.
151. Jemal, A., et al., *Cancer burden in Africa and opportunities for prevention.* Cancer, 2012. **118**(18): p. 4372-84.
152. Semeere, A., et al., *A prospective ascertainment of cancer incidence in sub-Saharan Africa: The case of Kaposi sarcoma.* Cancer Med, 2016. **5**(5): p. 914-28.
153. Asiimwe, S.B., et al. *Impact of Kaposi's sarcoma on survival in HIV-infected African adults on antiretroviral therapy*. in *21st Conference on Retroviruses and Opportunistic Infections*. 2014.
154. Malope, B.I., et al., *No evidence of sexual transmission of Kaposi's sarcoma herpes virus in a heterosexual South African population.* AIDS, 2008. **22**(4): p. 519-526.
155. Mayama, S., et al., *Prevalence and transmission of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in Ugandan children and adolescents.* Int J Cancer, 1998. **77**(6): p. 817-20.
156. Bourboulia, D., et al., *Serologic evidence for mother-to-child transmission of Kaposi sarcoma-associated herpesvirus infection.* JAMA, 1998. **280**(1): p. 31-2.
157. Etta, E.M., et al., *HHV-8 Seroprevalence and Genotype Distribution in Africa, 1998(-)2017: A Systematic Review.* Viruses, 2018. **10**(9).
158. Dollard, S.C., et al., *Substantial regional differences in human herpesvirus 8 seroprevalence in sub-Saharan Africa: insights on the origin of the "Kaposi's sarcoma belt".* Int J Cancer, 2010. **127**(10): p. 2395-401.
159. Olp, L.N., et al., *Effects of Antiretroviral Therapy on Kaposi's Sarcoma-Associated Herpesvirus (KSHV) Transmission Among HIV-Infected Zambian Children.* J Natl Cancer Inst, 2015. **107**(10).
160. Dedicoat, M., et al., *Mother-to-child transmission of human herpesvirus-8 in South Africa.* J Infect Dis, 2004. **190**(6): p. 1068-75.
161. Plancoulaine, S., et al., *Respective roles of serological status and blood specific antihuman herpesvirus 8 antibody levels in human herpesvirus 8 intrafamilial transmission in a highly endemic area.* Cancer Res, 2004. **64**(23): p. 8782-7.
162. Minhas, V., et al., *Early Childhood Infection by Human Herpesvirus 8 in Zambia and the Role of Human Immunodeficiency Virus Type 1 Coinfection in a Highly Endemic Area.* American Journal of Epidemiology, 2008. **168**(3): p. 311-320.
163. Butler, L.M., et al., *Human herpesvirus 8 infection in children and adults in a population-based study in rural Uganda.* J Infect Dis, 2011. **203**(5): p. 625-34.
164. Pfeiffer, R.M., et al., *Geographic heterogeneity of prevalence of the human herpesvirus 8 in sub-Saharan Africa: clues about etiology.* Ann Epidemiol, 2010. **20**(12): p. 958-63.

165. Mbondji-Wonje, C., et al., *Seroprevalence of human herpesvirus-8 in HIV-1 infected and uninfected individuals in Cameroon.* Viruses, 2013. **5**(9): p. 2253-9.
166. Jacky, N.B., et al., *[Seroprevalence of human herpes virus-8 in HIV-positive patients at the General Hospital of Yaounde - Cameroon].* Pan Afr Med J, 2015. **20**: p. 69.
167. Stolka, K., et al., *Risk factors for Kaposi's sarcoma among HIV-positive individuals in a case control study in Cameroon.* Cancer Epidemiol, 2014. **38**(2): p. 137-43.
168. Wilkinson, D., et al., *Prevalence of infection with human herpesvirus 8/Kaposi's sarcoma herpesvirus in rural South Africa.* S Afr Med J, 1999. **89**(5): p. 554-7.
169. Sitas, F., et al., *Antibodies against human herpesvirus 8 in black South African patients with cancer.* N Engl J Med, 1999. **340**(24): p. 1863-71.
170. Whitby, D., et al., *Genotypic characterization of Kaposi's sarcoma-associated herpesvirus in asymptomatic infected subjects from isolated populations.* J Gen Virol, 2004. **85**(Pt 1): p. 155-163.
171. Sallah, N., et al., *Genome-Wide Sequence Analysis of Kaposi Sarcoma-Associated Herpesvirus Shows Diversification Driven by Recombination.* J Infect Dis, 2018. **218**(11): p. 1700-1710.
172. Lacoste, V., et al., *Molecular epidemiology of human herpesvirus 8 in africa: both B and A5 K1 genotypes, as well as the M and P genotypes of K14.1/K15 loci, are frequent and widespread.* Virology, 2000. **278**(1): p. 60-74.
173. Alagiozoglou, L., F. Sitas, and L. Morris, *Phylogenetic analysis of human herpesvirus-8 in South Africa and identification of a novel subgroup.* J Gen Virol, 2000. **81**(Pt 8): p. 2029-2038.
174. Arvanitakis, L., et al., *Establishment and characterization of a primary effusion (body cavity-based) lymphoma cell line (BC-3) harboring kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) in the absence of Epstein-Barr virus.* Blood, 1996. **88**(7): p. 2648-54.
175. Said, W., et al., *Kaposi's sarcoma-associated herpesvirus (KSHV or HHV8) in primary effusion lymphoma: ultrastructural demonstration of herpesvirus in lymphoma cells.* Blood, 1996. **87**(12): p. 4937-43.
176. Tamburro, K.M., et al., *Vironome of Kaposi sarcoma associated herpesvirus-inflammatory cytokine syndrome in an AIDS patient reveals co-infection of human herpesvirus 8 and human herpesvirus 6A.* Virology, 2012. **433**(1): p. 220-5.
177. Komanduri, K.V., et al., *The natural history and molecular heterogeneity of HIV-associated primary malignant lymphomatous effusions.* J Acquir Immune Defic Syndr Hum Retrovirol, 1996. **13**(3): p. 215-26.
178. Yakushko, Y., et al., *Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome contains a duplication of a long unique-region fragment within the terminal repeat region.* J Virol, 2011. **85**(9): p. 4612-7.
179. Brulois, K.F., et al., *Construction and manipulation of a new Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome clone.* J Virol, 2012. **86**(18): p. 9708-20.
180. Cannon, J.S., et al., *A new primary effusion lymphoma-derived cell line yields a highly infectious Kaposi's sarcoma herpesvirus-containing supernatant.* J Virol, 2000. **74**(21): p. 10187-93.
181. Engelbrecht, S., et al., *Detection of human herpes virus 8 DNA and sequence polymorphism in classical, epidemic, and iatrogenic Kaposi's sarcoma in South Africa.* J Med Virol, 1997. **52**(2): p. 168-72.

182. Zong, J.C., et al., *Strain variability among Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genomes: evidence that a large cohort of United States AIDS patients may have been infected by a single common isolate.* J Virol, 1997. **71**(3): p. 2505-11.
183. Treurnicht, F.K., et al., *HHV-8 subtypes in South Africa: identification of a case suggesting a novel B variant.* J Med Virol, 2002. **66**(2): p. 235-40.
184. Blumenthal, M.J., et al., *EPHA2 sequence variants are associated with susceptibility to Kaposi's sarcoma-associated herpesvirus infection and Kaposi's sarcoma prevalence in HIV-infected patients.* Cancer Epidemiol, 2018. **56**: p. 133-139.
185. Lambarey, H., et al., *SARS-CoV-2 Infection Is Associated with Uncontrolled HIV Viral Load in Non-Hospitalized HIV-Infected Patients from Gugulethu, South Africa.* Viruses, 2022. **14**(6).
186. Lesmes-Rodríguez, L.C., et al., *Previous exposure to common coronavirus HCoV-NL63 is associated with reduced COVID-19 severity in patients from Cape Town, South Africa.* Frontiers in Virology, 2023. **3**.
187. Cook, P.M., et al., *Variability and evolution of Kaposi's sarcoma-associated herpesvirus in Europe and Africa. International Collaborative Group.* AIDS, 1999. **13**(10): p. 1165-76.
188. Bushnell, B., Rood, J., Singer, E. *BBMerge – Accurate paired shotgun read merging via overlap*. [Short read aligner and other bioinformatic tools] 2017 [cited 2022-2023; BBTools]. Available from: https://sourceforge.net/projects/bbmap/.
189. Prjibelski, A., et al., *Using SPAdes De Novo Assembler.* Current Protocols in Bioinformatics, 2020. **70**(1): p. e102.
190. Prjibelski, A.D., et al., *ExSPAnder: a universal repeat resolver for DNA fragment assembly.* Bioinformatics, 2014. **30**(12): p. i293-i301.
191. Vasilinetc, I., et al., *Assembling short reads from jumping libraries with large insert sizes.* Bioinformatics, 2015. **31**(20): p. 3262-3268.
192. Johnson, M., et al., *NCBI BLAST: a better web interface.* Nucleic Acids Research, 2008. **36**(suppl_2): p. W5-W9.
193. Bosi, E., et al., *MeDuSa: a multi-draft based scaffolder.* Bioinformatics, 2015. **31**(15): p. 2443-2451.
194. Li, H., *Minimap2: pairwise alignment for nucleotide sequences.* Bioinformatics, 2018. **34**(18): p. 3094-3100.
195. Li, H., *New strategies to improve minimap2 alignment accuracy.* Bioinformatics, 2021. **37**(23): p. 4572-4574.
196. Katoh, K., et al., *MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.* Nucleic Acids Research, 2002. **30**(14): p. 3059-3066.
197. Katoh, K. and D.M. Standley, *MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.* Molecular Biology and Evolution, 2013. **30**(4): p. 772-780.
198. Minh, B.Q., et al., *IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.* Molecular Biology and Evolution, 2020. **37**(5): p. 1530-1534.
199. Hoang, D.T., et al., *UFBoot2: Improving the Ultrafast Bootstrap Approximation.* Molecular Biology and Evolution, 2017. **35**(2): p. 518-522.
200. Kalyaanamoorthy, S., et al., *ModelFinder: fast model selection for accurate phylogenetic estimates.* Nature Methods, 2017. **14**(6): p. 587-589.

201. Raumbaut, A. *FigTree v1.4.4*. [Software] 2018 [cited 2022-2023; Phylogenetic visualizer]. Available from: http://tree.bio.ed.ac.uk/software/figtree/.
202. Huson, D.H., *SplitsTree: analyzing and visualizing evolutionary data.* Bioinformatics, 1998. **14**(1): p. 68-73.
203. Awazawa, R., et al., *High Prevalence of Distinct Human Herpesvirus 8 Contributes to the High Incidence of Non-acquired Immune Deficiency Syndrome-Associated Kaposi's Sarcoma in Isolated Japanese Islands.* J Infect Dis, 2017. **216**(7): p. 850-858.
204. Santiago, J.C., et al., *Tumor-specific changes in Kaposi sarcoma-associated herpesvirus genomes in Ugandan adults with Kaposi sarcoma.* bioRxiv, 2020: p. 2020.05.04.076638.
205. Moorad, R., et al., *Whole-genome sequencing of Kaposi sarcoma-associated herpesvirus (KSHV_HHV8) reveals evidence for two African lineages.* Virology, 2022. **568**: p. 101-114.
206. Meng, Y.X., et al., *Molecular characterization of strains of Human herpesvirus 8 from Japan, Argentina and Kuwait.* J Gen Virol, 2001. **82**(Pt 3): p. 499-506.
207. Fouchard, N., et al., *Detection and genetic polymorphism of human herpes virus type 8 in endemic or epidemic Kaposi's sarcoma from West and Central Africa, and South America.* Int J Cancer, 2000. **85**(2): p. 166-70.
208. Meng, Y.X., et al., *Individuals from North America, Australasia, and Africa are infected with four different genotypes of human herpesvirus 8.* Virology, 1999. **261**(1): p. 106-19.
209. Duprez, R., et al., *Molecular epidemiology of the HHV-8 K1 gene from Moroccan patients with Kaposi's sarcoma.* Virology, 2006. **353**(1): p. 121-32.
210. Tozetto-Mendoza, T.R., et al., *Genotypic distribution of HHV-8 in AIDS individuals without and with Kaposi sarcoma: Is genotype B associated with better prognosis of AIDS-KS?* Medicine, 2016. **95**(48): p. e5291.
211. Kakoola, D.N., et al., *Recombination in human herpesvirus-8 strains from Uganda and evolution of the K15 gene.* J Gen Virol, 2001. **82**(Pt 10): p. 2393-2404.
212. Caro-Vegas, C., et al., *Runaway Kaposi Sarcoma-associated herpesvirus replication correlates with systemic IL-10 levels.* Virology, 2020. **539**: p. 18-25.
213. Taylor, J.L., et al., *Transcriptional analysis of latent and inducible Kaposi's sarcoma-associated herpesvirus transcripts in the K4 to K7 region.* J Virol, 2005. **79**(24): p. 15099-106.
214. Feng, P., et al., *Kaposi's sarcoma-associated herpesvirus mitochondrial K7 protein targets a cellular calcium-modulating cyclophilin ligand to modulate intracellular calcium concentration and inhibit apoptosis.* J Virol, 2002. **76**(22): p. 11491-504.
215. Wang, Y., et al., *Functional Characterization of Kaposi's Sarcoma-Associated Herpesvirus Open Reading Frame K8 by Bacterial Artificial Chromosome-Based Mutagenesis.* Journal of Virology, 2011. **85**(5): p. 1943-1957.
216. Liu, D., Y. Wang, and Y. Yuan, *Kaposi's Sarcoma-Associated Herpesvirus K8 Is an RNA Binding Protein That Regulates Viral DNA Replication in Coordination with a Noncoding RNA.* J Virol, 2018. **92**(7).
217. Chauhan, R.P., et al., *Systematic Review of Important Viral Diseases in Africa in Light of the 'One Health' Concept.* Pathogens, 2020. **9**(4).
218. Suzane, et al., *KSHV genotypes A and C are more frequent in Kaposi sarcoma lesions from Brazilian patients with and without HIV infection, respectively.* Cancer Letters, 2011. **301**(1): p. 85-94.

219. Nalwoga, A., et al., *Comparison of Epstein-Barr virus and Kaposi's sarcoma-associated herpesvirus viral load in peripheral blood mononuclear cells and oral fluids of HIV-negative individuals aged 3 to 89 years from Uganda.* Res Sq, 2023.
220. Nalwoga, A., et al., *Risk Factors for Kaposi's Sarcoma-Associated Herpesvirus DNA in Blood and in Saliva in Rural Uganda.* Clin Infect Dis, 2020. **71**(4): p. 1055-1062.

8. Supplementary Material

Table 8: Information of all samples selected for this study. Shown are the sample name, original cohort, viral load, DNA concentration, available volume, as well as K1 and K15 subtypes. Greyed out blocks indicate that no sequence is available.

#	Name	Cohort	Viral Load (Raw KSHV copies/10 µL)	DNA conc. (ng/µL)	Approx. volume remaining (µL)	K1	K15
1	KDHTB153	KDHTB	2136.67	124.0	24.2	A5	N
2	KDHTB658	KDHTB	88033.33	86.8	34.6	A5	N
3	KDHTB216	KDHTB	1533.33	12.1	12.0	B3	P
4	KDHTB268	KDHTB	681.67	34.4	175.0	B2	P
5	KDHTB008	KDHTB	378.33	39.1	50.0	B2	N
6	KDHTB566	KDHTB	75.60	65.6	105.0	A5	
7	KDHTB496	KDHTB	61.80	42.1	120.0	A5	
8	KDHTB296	KDHTB	18.40	34.3	150.0	A5	
9	KDHTB010	KDHTB	16.20	39.1	35.0	A5	
10	KDHTB096	KDHTB	11.45	42.7	65.0	A5	
11	KDHTB528	KDHTB	5535.00	94.9	25.0		
12	KDHTB277	KDHTB	175.67	36.1	200.0		
13	KDHTB095	KDHTB	34.53	82.4	120.0		
14	KDHTB223	KDHTB	33.60	35.5	100.0		
15	KDHTB094	KDHTB	26.63	53.3	75.0		
16	KDHTB329	KDHTB	16.11	44.3	60.0		
17	KDHTB362	KDHTB	12.26	20.3	150.0		
18	KDHTB676	KDHTB	5.77	25.6	80.0		
19	KDHTB386	KDHTB	3.94	56.8	100.0		
20	KDHTB299	KDHTB	3.78	9.4	130.0		
21	KS029	KS	1090.00	65.5	50.0	A5	M
22	KS041	KS	2103.33	375.3	52.4	A5	N
23	MB033	KS	3766.00	38.7	25.0	A5	M
24	KS036	KS	161.00	240.1	20.0	A5	
25	KS059	KS	202.67	158.3	100.0	A5	
26	KS078	KS	118.00	24.0	30.0	A5	
27	KS010	KS	79.59	46.3	35.0	A5	
28	KS065	KS	38.40	94.7	85.0	A5	
29	KS062	KS	36.80	65.6	60.0	A5	
30	KS040	KS	36.77	194.0	65.0	A5	
31	KS054	KS	33.60	20.3	50.0	A5	
32	KS088	KS	29.20	89.9	75.0	A5	
33	KS085	KS	21.60	5.8	50.0	A5	
34	KS094	KS	10.62	36.8	60.0	A5	
35	KS080	KS	6.62	79.2	70.0	A5	

36	KS099	KS	5.29	62.4	60.0	A5	
37	KS007	KS	4.37	38.4	60.0	A5	
38	KS035	KS	108.00	153.3	12.0		
39	KS001	KS	3513.33	98.9	0.0		
40	KS028	KS	960.00	58.5	0.0		
41	KS20	KS	141.67	89.6	50.0		
42	KS021	KS	93.67	67.6	40.0		
43	KS074	KS	81.63	89.5	60.0		
44	KS082	KS	67.23	125.3	75.0		
45	KS081	KS	56.90	286.7	80.0		
46	KS069	KS	43.17	36.1	125.0		
47	KS039	KS	34.93	144.3	60.0		
48	KS042	KS	30.00	363.3	45.0		
49	KS016	KS	10.34	154.0	60.0		
50	KS098	KS	9.99	33.9	55.0		
51	KS003	KS	8.43	30.0	40.0		
52	KS101	KS	6.31	26.4	45.0		
53	KS086	KS	5.94	207.0	75.0		
54	KS023	KS	3.36	78.4	40.0		
55	GUG116	GUG	244.30	114.0	70.0	A5	M
56	GUG069	GUG	13.31	73.2	90.0	A5	
57	GUG064	GUG	31.47	39.9	110.0		

Table 9: Table of reference-guided alignments statistics. The statistics of the reference-guided alignment sequences created through our collaborators pipeline and used to compare this study's de novo pipeline quality. The same sequencing data was used. KDHTB528 failed next generation sequencing and has been greyed out.

ID	Raw Reads	Total number of reads mapped to reference	Number of bases of reference covered	Average Coverage	Median Coverage	Per-base coverage range, zero to most-covered	% of bases w/ coverage greater than 10x	% of reference bases covered	Sample Comments
KDHTB153_SA	255749	155749	137269	198.418	203	0:366	0.991526	99.60	
KDHTB658_SA	461592	149882	137009	191.092	198	0:296	0.992589	100.00	
KDHTB216_SA	4115860	142817	137494	189.928	197	0:327	0.99169	99.96	
KDHTB268_SA	2026054	2015675	137354	1793.94	1753	0:10934	0.99611	99.70	
KDHTB008_SA	428404	426168	136246	378.977	340	0:3410	0.988314	98.90	
KDHTB528	3731128	306	1213	3.38029	0	0:90	0.0637858	1.00	Failed sequencing
KS029_SA	335956	328525	134808	297.006	274	0:2997	0.992935	97.80	
KS041_SA	1852186	149558	136630	190.961	199	0:294	0.987821	99.50	
MB033_SA	12066838	11872914	135454	10570.3	10142	0:113298	0.998088	98.20	
GUG116_SA	1951338	1909024	134986	1713.47	1695	0:10293	0.990487	98.20	