

## Whole-genome sequencing of Kaposi sarcoma-associated herpesvirus (KSHV/HHV8) reveals evidence for two African lineages

Razia Moorad, Angelica Juarez, Justin T. Landis, Linda J. Pluta, Megan Perkins, Avery Cheves, Dirk P. Dittmer\*

Lieberger Comprehensive Cancer Center, School of Medicine, Department of Immunology and Microbiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

### ARTICLE INFO

**Keywords:**  
Herpesvirus  
Kaposi sarcoma  
Genome  
AmpliSeq  
Ion torrent  
KSHV

### ABSTRACT

Kaposi sarcoma (KS)-associated herpesvirus (KSHV/HHV-8) was first sequenced from the body cavity (BC) lymphoma cell line, BC-1, in 1996. Few other KSHV genomes have been reported. Our knowledge of sequence variation for this virus remains spotty. This study reports additional genomes from historical US patient samples and from African KS biopsies. It describes an assay that spans regions of the virus that cannot be covered by short read sequencing. These include the terminal repeats, the LANA repeats, and the origins of replication. A phylogenetic analysis, based on 107 genomes, identified three distinct clades; one containing isolates from USA/Europe/Japan collected in the 1990s and two of Sub-Saharan Africa isolates collected since 2010. This analysis indicates that the KSHV strains circulating today differ from the isolates collected at the height of the AIDS epidemic. This analysis helps experimental designs and potential vaccine studies.

### 1. Introduction

Kaposi Sarcoma-associated Herpesvirus (KSHV) or human herpesvirus 8 (HHV8) is the etiological agent of Kaposi's sarcoma (KS), Primary Effusion Lymphoma (PEL), and the plasmablastic variant of Multicentric Castlemann's Disease (MCD) (reviewed in (Dittmer and Damania, 2016)). It also causes acute disease such as KS-Immune Reconstitution Inflammatory Syndrome (KS-IRIS) and KSHV Inflammatory Cytokine Syndrome (KICS). KS is the most common cancer in persons living with HIV (PLWH), today as well as in PLWH who progressed to acquired immunodeficiency syndrome (AIDS). After an initial decline due to the widespread availability of combination antiretroviral therapy (cART), KS cases have stabilized across the world. In some locales KS cases have started to increase as the cohort of initially infected PLWH ages and as the rate of new HIV infections no longer declines. Since 2005 about one third of KS cases develop in PLWH on stable cART, with suppressed HIV viral loads and near normal CD4 counts (Krown et al., 2008; Maurer et al., 2007; Royston et al., 2021). In those patients KS is uncoupled from HIV replication and HIV-induced CD4 cell depletion. As HIV infections continue to remain a public health problem indefinitely, so will KSHV-associated malignancies.

KS is an angioproliferative, cytokine-driven neoplasm, similar to

angiosarcoma (Schneider and Dittmer, 2017). HIV KS presents mainly as skin lesions but also affects mucosal membranes and visceral organs. In endemic regions pediatric KS may manifest exclusively within lymph nodes in the absence of skin lesions (El-Mallawany et al., 2018, 2019b). Infections with KSHV may also give rise to KICS and KS-IRIS. The latter is frequent upon initiation of cART in KSHV endemic regions (Nyirenda et al., 2020). PEL is a rare and aggressive non-Hodgkin's B-cell lymphoma also known as Body Cavity-based Lymphoma (BCBL); like KS it is considered an AIDS-defining malignancy. PEL has a poor prognosis with a mean survival time of 2–6 months and may develop in conjunction with KS. On rare occasions, PEL has been described in HIV-negative, KSHV-positive persons with highly suppressed immune systems (Jones et al., 1998; Klein et al., 2003). As PEL cells are highly transformed they can be adapted to continued growth in culture. PEL growth is dependent on KSHV, and in some cases also on co-infection with Epstein-Barr Virus (EBV) (Bigi et al., 2018; Godfrey et al., 2005). PEL cell lines are the most widely used model for studying KSHV pathogenesis.

Like other herpesviruses, the KSHV life cycle consists of two alternating forms of infection: latent and lytic. Latency is considered the default pathway for KSHV and is observed in most KS lesions and virtually all tissues, including cell lines of endothelial, lymphoid, epithelial, and mesenchymal origin (Hosseini-pour et al., 2014; Jones

\* Corresponding author.

E-mail address: [ddittmer@med.unc.edu](mailto:ddittmer@med.unc.edu) (D.P. Dittmer).

<https://doi.org/10.1016/j.virol.2022.01.005>

Received 2 October 2021; Received in revised form 7 January 2022; Accepted 16 January 2022

Available online 2 February 2022

0042-6822/© 2022 Elsevier Inc. All rights reserved.

et al., 2012; Renne et al., 1998). During latency, the KSHV DNA exists as nonintegrated nuclear episome or plasmid. Only a limited number of viral genes are expressed, including the latency-associated nuclear antigen (LANA; ORF73) (Dittmer et al., 1998; Kedes et al., 1997; Rainbow et al., 1997), which is used as the defining clinical diagnostic marker for KS and PEL. Overall, the KSHV genome encodes over 80 ORFs, several long non-coding RNAs and microRNAs (miRNAs), although the actual number of transcripts and peptides is likely much higher (Arias et al., 2014; Bai et al., 2014; Dittmer, 2003; Dresang et al., 2011; Majerciak et al., 2013; Sarid et al., 1998). KSHV establishes systemic, lifelong latency primarily in B-lymphocytes. Unlike EBV, the KSHV latent reservoir seems tissue resident; few KSHV positive B cells circulate during asymptomatic latency (Decker et al., 1996) and the viral genome is barely detectable in plasma or peripheral blood mononuclear cells (PBMC). This has made it difficult to study the virus.

Much effort has been expended to uncover the biological and functional details of KSHV to develop better diagnostic options, potential vaccines, and novel treatment regimens. These efforts are hampered by the lack of experimental models. KSHV does not productively infect any mammal other than humans. Phenotypes in KSHV exposed non-human primates are minimal (Chang et al., 2009). Only humanized mice seem to support KSHV infection (Dittmer et al., 1999; McHugh et al., 2017; Wang et al., 2014). Notably, all infection models to date have used one strain of KSHV, which is derived from the JSC-1 PEL cell line. The JSC-1 isolate also represents the only instance of a clear plaque caused by KSHV (Cannon et al., 2000). In addition to the JSC-1 cell line most experiments use just a few other cell lines: BCBL-1, BC-3, and BC-1 (Cannon et al., 2000; Cesarman et al., 1995; Renne et al., 1996b). These cell lines were derived from HIV+ patients who succumbed to AIDS in the 1990s in the US. Thus, the vast majority of experimental KSHV research, including drug discovery and vaccine efforts, is based on three, 40-year-old virus isolates. This introduces an enormous liability as to robustness, reproducibility, and universality of KSHV research and prompted our study to estimate how related those selected few, initial KSHV isolates are to strains that are circulating today.

The KSHV genomic architecture is like that of other herpesviruses. It consists of a single long unique region (LUR) flanked by highly GC-rich terminal repeats (TR) at both termini of the genome (Lagunoff and Ganem, 1997; Moore et al., 1996; Nicholas et al., 1998; Renne et al., 1996a; Russo et al., 1996). The discovery of KSHV sparked early efforts to type and subtype KSHV. At the time, next generation sequencing had not been invented and efforts focused on highly polymorphic regions of the genome, such as the K1 gene or the LANA repeat region (Gao et al., 1999). The K1 gene proved useful as it is under constant immune selection. The extracellular region of the K1 gene, located at the 5' end of the viral genome, displays up to 30% sequence variability. This facilitated the classification of KSHV into six subtypes (A-F) (Cook et al., 1999; Meng et al., 1999). These early studies informed our knowledge about the geographic distribution of KSHV (Hayward, 1999; Meng et al., 1999), but were limited by technology and unable to go beyond a single gene as substitute marker for the evolution of the entire viral genome.

Recent studies have begun to fill this gap in our knowledge, particularly regarding strains that circulate in endemic regions, i.e., Sub-Saharan Africa (SSA). We reported the first complete KSHV sequence from cell free virus from the blood of a KICS patient (Tamburro et al., 2012) and more recently a second individual (Caro-Vegas et al., 2020); both from the US. Olp et al. reported sequences from KS biopsies in Zambia (Olp et al., 2015), Santiago et al. from KS biopsies and saliva in Uganda (Santiago et al., 2021) and Sallah et al. from saliva from asymptomatic KSHV carriers in Uganda (Sallah et al., 2018). Here, we report additional sequences from KS patients in Malawi as well as additional samples from the US. While these data are still the result of clinical opportunities, rather than of carefully designed population-wide studies, they reveal the existence of two distinct African lineages of KSHV that are not represented in current experimental studies. This novel information should be useful for KSHV vaccine designs as well as the

interpretation of current experimental results.

To facilitate further studies, we also report a set of validated PCR primers that amplify regions of KSHV that are not typically reported in next generation sequencing studies, such as the LANA repeats, the TRs, and the viral origins of replication and other low complexity regions. These should be useful to assess the stability of experimental models such as KSHV recombinant bacmids or long-term cultured cell lines (Brulois et al., 2012; Budt et al., 2011; Jain et al., 2016; Yakushko et al., 2011; Zhou et al., 2002).

## 2. Methods

**Samples.** PEL patient samples were obtained from the AIDS and Cancer Specimen Resource (ACSR). A total of seven KSHV PEL samples including 30039264 (pleura-PEL cells), 30039265 (Esophagus, PEL cells), 30039266 (lung, FFPE curls), 30039267 (stomach, PEL cells), 30007535 (ascites, PEL cells), 30007536 (ascites, PEL cells) and 30007537 (ascites, PEL cells). Three additional samples were received as normal samples and include 30035668 (Lymph node, cells), 30039263 (skin, cells), and 3004479 (spleen, cells). These were received as formalin-fixed paraffin embedded (FFPE) curls or as frozen cell pellets as indicated. All samples were also residual pathology specimen. Their use was approved by the institutional review board (IRB) of the University of North Carolina (No.:09-1201, 14-0221, 14-2151).

**Sample DNA preparation.** For FFPE, DNA was extracted using the QIAamp DNA FFPE Tissue Kit, following the manufacturer's instructions. For other types of samples, DNA was extracted using the Roche MagNA Pure Compact Instrument (Roche, #0373114601). 10  $\mu$ l of a control plasmid, Fly2.0 (Addgene #117418) ( $1.0 \times 10^5$  copies/ $\mu$ l), was added to each sample to serve as a control for DNA extraction efficiency.

**Library preparation, target enrichment, and Ion Torrent sequencing.** Total DNA was quantified by Qubit 3.0 dsDNA HS Assay (Life Technologies). Custom-made Ion AmpliSeq primer pools (Life Technologies) were designed to amplify the viral genome using KSHV BAC16 (JSC1 isolate, GenBank accession number GQ994935) as the target. Libraries were manually prepared from 100 ng total DNA using the Ion AmpliSeq Library Preparation Kit 2.0 (Life Technologies), protocol MAN00006735 Rev: F.0. Libraries were quantified and sized with Agilent Bioanalyzer 2100 High Sensitivity DNA Assay (Agilent Technologies) and pooled to 100pM final concentration. Templating and loading onto the Ion 530 Chip (Life Technologies) were automated on the Ion Chef (Life Technologies). Samples were sequenced on the Ion Torrent S5 (Life Technologies) with default parameters. Ion Torrent barcodes and adapter sequences are removed prior to output of FASTA sequence data files from the Ion Torrent S5 server.

**Quality trimming, filtering, and read mapping.** Reads were trimmed by low-quality base pairs (quality limit = 0.05), including 40 nucleotides at the 5' terminal and 10 nucleotides at the 3' terminal. All reads shorter than 50 nucleotides were filtered out. High-quality, trimmed reads were mapped to the KSHV genome (GenBank accession number NC\_009333) using the map-to-reference tool on CLC Genomics Workbench v20.0.3 (Qiagen) with default parameters. A length fraction of 0.5 and a similarity fraction of 0.8 were selected and non-specific reads were ignored for better mapping accuracy. Duplicate mapped reads were removed using default parameters on the remove-duplicate-mapped-reads-tool on CLC Genomics Workbench v20.0.3.

**Variant Calling.** Variants including Single Nucleotide Variants (SNVs), Multiple Nucleotide Variants (MNVs), and Insertions and Deletions (InDels) were called using the Basic Variant Detection tool on CLC Genomics Workbench v20.0.3 (Qiagen), using default parameters. SNVs with a minimum average Phred quality score >20, a minimum frequency >60%, a minimum coverage >40, and a minimum forward/reverse balance >0.05, were reported.

**Consensus Sequence.** Consensus sequences were built on CLC Genomics Workbench (v20.0.3) based on quality score voting, with low

coverage regions being defined as regions with a minimum of 3 reads and a maximum of the total number of reads. Low coverage regions were filled from the reference sequence (NC\_009333) and annotated as such. All complete sequences have been submitted to GenBank.

**Sequencing primers, PCR amplification, and cloning of KSHV-BAC16 Ion AmpliSeq amplicon gap regions.** Primers (Sigma-Aldrich) were designed using Benchling (Benchling Inc.). Predicted T<sub>m</sub> values and other parameters were evaluated using the Sigma-Aldrich online tools. Gradient PCR was performed with each primer pair using between 10 and 500 ng of KSHV-BAC DNA or 2-5 ng of the sample DNA was amplified in a 25 µl gradient PCR reaction using OneTaq Hot Start Polymerase kit (New England Biolabs, Cat. #M0480). PCR conditions included an initial denaturation step for 30 s, followed by PCR cycling at 94 °C for 15 s, 55-65 °C for 30 s and 75 °C for 2 min for 30 cycles (increased 1 min/kb for the larger gap regions), followed by a final extension at 75 °C for 10 min, to ensure 3' overhangs. The primer pairs and PCR conditions are summarized in Table 4. PCR products were gel purified using Qiagen Gel Purification Kit (Qiagen Inc.) and ligated into the pCR4-TOPO vector (Invitrogen Inc.). Positive clones were identified by restriction digest and plasmid DNA subjected to Sanger Sequencing. Alternatively, tagging KSHV-specific primers with the universal M13 primers, allowed for direct Sanger sequencing of the PCR product. Samples were purified using the Qiagen PCR purification kit (Qiagen Inc.) and purified DNA subjected to Sanger Sequencing.

**Phylogenetic analysis.** Isolates were classified according to their K1 genotypes, by aligning all consensus sequences generated in this study with 131 previously published representative K1 genotypes (Supplementary Table 1). K1 orf sequences were aligned using MAFFT implemented in Geneious v9.1.8 with default parameters (Katoh and Standley, 2013). The translated multiple alignments were used to infer the maximum likelihood (ML) tree using the General Time Reversible (GTR) protein model with the gamma distribution for site variation, implemented in RAxML in Geneious v9.1.8 (Stamatakis, 2014). Bootstrap statistical analysis of 1000 replicates was performed. The K15 genotyping was performed similarly using representative K15 genotypes of the P-allele (AF148805.2) (Glenn et al., 1999), the M-allele (AF156885.1) (Nicholas et al., 1998) and the N-allele (Olp et al., 2015).

The KSHV isolates described in this study were compared to 108 publicly available whole KSHV genomes from GenBank (Table 3), including 23 Ugandan KSHV isolates, for which SRA data were publicly available (Sallah et al., 2018). The whole genomes of all isolates were aligned using the MAFFT with default parameters implemented in Geneious v9.1.8. The ML tree was built as described above and trees were rooted to BC-1 (MK733607). The ML tree on the whole genome included only the LUR region between the two Ori-Lyt sites in KSHV (K7-ORF58). This yielded 67,021bp of contiguous sequences, which was extracted from the alignment. The evolutionary relationship of all African isolates and BAC sequences were analyzed using the same methodology described above, with an alignment of the central regions of

KSHV extracted from the BAC sequences and the African isolates listed in Table 2. The MAFFT alignment of the KSHV-LUR was used to construct the Bayesian time-dated phylogenetic tree with BEAST v1.10.4 (Drummond and Rambaut, 2007; Drummond et al., 2012), using the HKY substitution model implemented in BEAST. The dates for the genomes described in this study were inferred from the collection dates and dates for the genomes retrieved from GenBank were inferred from the submission dates. A gamma site model was used to estimate site substitution rates. A strict clock and a constant size coalescent restriction were used to construct the tree.

## 2.1. Recombination analysis

The Splits Tree analysis was performed to investigate conflicting phylogenetic signals and was performed on the ML tree alignment of the LUR, using SplitsTree4 (Huson and Bryant, 2006). The Neighbor-net split network was constructed, using uncorrected P characters transformation, and excluding gap sites. 1000 bootstrap replicates and the Phi test for evidence of recombination were performed. Possible breakpoints were investigated using five different methods implemented in the Recombination Detection Program 4 (RDP4; v.4.101). These include the RDP, BootScan, GENECONV, MaxChi and SisScan. A window size of 2,000 and a step size of 200, was used to determine recombination; evidence of statistically significant genetic recombination was confirmed by the concurrence of two or more methods in RDP4. The similarity or relatedness of the African and USA genomes described in this study to a set of reference and KSHV-BAC genomes was investigated using SimPlot (v3.5.1). The Neighbor-joining method based on the Kimura two-parameter model, with 1,000 bootstrap replicates, and a window size of 2000 bp and a step size of 20 bp was used to compute similarity.

## 3. Results

### 3.1. Established cell lines and bacmids exhibit limited variation in KSHV

The first KSHV genome was determined by Sanger sequencing, using phage and cosmid libraries from the BC-1 PEL cell line (Russo et al., 1996). The coverage was ~ 6x (GenBank ID: KSU75698 or U75698). We re-sequenced BC-1 in 2019 using short read technology (GenBank ID: MK733607). The current, KSHV GenBank reference sequence was obtained from a KS biopsy of European origin, also determined by Sanger sequencing (GenBank ID: NC\_009333, identical to AF148805). It is referred to as the GK18 strain. The BC-1 cell line gave rise to bacmid BAC-36 (Yakushko et al., 2011; Zhou et al., 2002) (GenBank ID: HQ404500) and various targeted mutant bacmid derivatives (GenBank ID: JX228174 (Bala et al., 2012)). The full-length BAC construct, KSHV-BAC16 is derived from the rKSHV.219 virus in JSC-1 PEL cells (GenBank ID: GQ994935.1) (Brulois et al., 2012). The rKSHV.219 was

**Table 1**

Metadata and mapping results of the 10 PEL samples obtained from the ACSR. K1 and K15 genotypes are listed for those genomes that had coverage in the respective regions. Genbank accession numbers for high coverage genomes are indicated.

ID/GenBank	Site/Pathology	Age (1)	Race	No. Total reads	No. mapped reads (2)	Mean coverage	K1/K15 genotype
1/MZ712173	Parietal pleura/NHL	32	White	1,304,839	478,390	x 464	P/-
2/MZ712174	Esophagus/NHL	31	White	1,250,899	512,015	x 545	C3/P
3	Lung/NHL	48	African American	595,453	179,737	x 127	-
4	Stomach/NHL	53	African American	50	49	x 0.02	-
5	Ascites/NHL	41	Other	335	308	x 0.35	-
6	Ascites/NHL	53	Other	31	31	x 0.03	-
7	Ascites/NHL	33	White	22	22	x 0.02	-
8/MZ712175	Lymph node/normal	49	African American	393,710	164,209	x 154	C3/P
9	Skin/normal	32	White	241,598	111,365	x 119	-
10/MZ712176	Spleen/normal	31	White	1,564,239	598,376	x 594	C3/P

(1) All samples were from HIV+ men.

(2) Unique reads with duplicates removed and mapped to NC\_009333.

**Table 2**

Metadata and mapping results of the archival KS samples from Malawi and the USA. Samples isolated from the USA were KS saliva samples. K1 and K15 genotypes are listed for those genomes that had coverage in the respective regions. Genbank accession numbers for high coverage genomes are indicated.

ID/GenBank	Type (1)	Region	No. Total reads	No. mapped reads (2)	Mean coverage	K1/K15 genotype
1	biopsy	Malawi	13,672	9231	x 13	–
2	biopsy	Malawi	13,295	9811	x 14	–
3/MZ712178	biopsy	Malawi	4,914,663	373,850	x 512	–/P
4/MZ712179	biopsy	Malawi	1,599,518	217,143	x 214	A5/-
5/MZ712182	biopsy	Malawi	1,474,915	209,444	x 262	A5/-
6	biopsy	Malawi	84	78	x 0.1	–
7/MZ712177	biopsy	Malawi	4,364,672	312,247	x 405	A5/-
8	biopsy	Malawi	29,061	24, 850	x 22	–
9 (3)	biopsy	Malawi	1174	1042	x 2.5	–
10/MZ712180	biopsy	Malawi	1,327,445	249,047	x 314	P/-
11/MZ712181	saliva	USA	6,777,102	331,951	x 486	C3/P
12/MZ712183	saliva	USA	2,185,672	490,827	x 700	C3/P
13/MZ712184	saliva	USA	3,759,465	406,242	x 582	A/-
14/MZ712185	saliva	USA	7,583,500	566,490	x 817	A/P

(1) All samples were from HIV+ participants with clinical KS.

(2) Unique reads with duplicates removed and mapped to NC\_009333.

(3) Shotgun sequencing.

sequenced independently (GenBank ID: KF588566) (Kati et al., 2015); and we and others have since sequenced this bacmid and various mutant derivatives (GenBank ID: MK733609, MN752405, KX189629, KX189628, KX189627, KX189626) (Beauclair et al., 2020; Zhang et al., 2016). The KSHV strain in the BCBL-1 cell line (GenBank ID: MN205539), as well as the KSHV strain of the BC-3 cell line (GenBank ID: MK876731) have also been sequenced by us and others. The origin of this data is summarized in Table 3

A comparison of these sequences allows for the investigation of the sequence variability in clonal, long-term propagated PEL cell lines. Here, one can assume that the viral genome is propagated by the error-correcting human DNA-dependent DNA polymerase, as 99% of cells in culture are latently infected. If the PEL cell lines were properly maintained under optimal growth condition one can further assume that selection is minimal. In the first approximation, except for the well-established propensity of the viral TR to contract (Ballestas et al., 1999; Hu et al., 2002), no further rearrangements should be induced, selected for, or stably maintained. Likewise, a comparison of the different bacmid sequences should allow for the estimation of sequencing errors, assuming, again that the *E. coli* DNA-dependent DNA polymerase is error correcting. Bacmid propagation of KSHV is known to result in the rapid loss of TR sequences, particularly under suboptimal growth conditions but except under selection against DNA motifs that are toxic in *E. coli*, so-called “poison sequences”, no selection pressure should be exerted on viral genes.

In the case of bacmids all sequences originate from the single seed clone as the common ancestor to all BAC36 and all BAC16 sequences. In case of PEL cell lines, we also assume a single common ancestor due to the bottleneck of converting a primary explant into a cell-culture adapted cell line, although none of the current PEL cell lines were single cell cloned. Indeed, an alignment of the bacmid sequences referenced in Table 3, confirmed their conserved DNA sequences. No obvious genomic re-arrangements were observed, except the previously reported duplication in BAC36 (Yakushko et al., 2011).

### 3.2. Novel KSHV genomes derived from primary PEL biopsies, KS biopsies and saliva

Until recently, few KSHV isolates have been sequenced from patient biopsies or blood (Caro-Vegas et al., 2020; Jary et al., 2020; Olp et al., 2015; Sallah et al., 2018; Tamburro et al., 2012). Recently, several studies have reported KSHV genomes from patient samples isolated from regions in SSA including Zambia and Uganda (Kajumbula et al., 2006; Olp et al., 2015; Sallah et al., 2018; Santiago et al., 2021).

By comparison to EBV, which has more than 1,000 genomes

sequenced (Xu et al., 2019), fewer than 100 complete genomes exist for KSHV. This limits any analyses of virus diversity and evolution and represent a barrier to scientific progress.

Firstly, this study contributes KSHV genomes from three primary PEL patient samples obtained from the AIDS and Cancer Specimen Resource (ACSR) repository. PEL sample UNC\_RM\_30039264 yielded 1,305,049 reads that aligned to the KSHV reference genome, resulting in a mean coverage of  $x475 \pm 873$ -fold. Sample UNC\_RM\_30039265 yielded 1,251,109 reads with a mean coverage of  $x551 \pm 628$ -fold and sample UNC\_RM\_30039266 yielded 595,699 reads with a mean coverage of  $420 \pm 3812$ -fold. Other archived material did not yield enough reads for reliable genome reconstitution (Table 1). Secondly, this study contributes KSHV genomes from two samples of histologically normal lymph node and spleen tissue from AIDS patients, most representing disseminated KSHV virions, micro metastases or circulating infected cells. Sample UNC\_RM\_30035668 yielded 393,710 reads (mean coverage of  $154 \pm 342$ -fold) and sample UNC\_RM\_30004479 yielded 1,564,239 reads (mean coverage  $594 \pm 1049$ -fold) respectively. Thirdly, this study contributes KSHV genomes from archival KS samples from Malawi and saliva from the US (Dittmer et al., 2017; Shiboski et al., 2015) resulting in an additional sixteen genomes. These are convenience samples that demonstrate the broad applicability of the AmpliSeq-technology based enrichment method to any type of long-time stored sample. AmpliSeq enrichment is PCR-based rather than bait-based. The metadata and mapping statistics are summarized in Table 2. Only genomes that obtained high coverage were included in further analysis. Their GenBank accession numbers are reported in Tables 1 and 2.

To validate the AmpliSeq primer set and to determine the effect of the DNA concentration on the reproducibility of consensus and variant determination the following experiment was conducted. Using BCBL-1 DNA automatic library preparation was validated using either the Ion Chef or the Genexus platform. 1 ng total DNA of input yielded a mean coverage of  $\sim 950$  fold  $\pm 300$ ,  $n = 4$  using the Ion Chef platform and a mean coverage of  $\sim 970$  fold  $\pm 150$ ,  $n = 4$  using the Genexus Platform, which is important as these, unlike manual methods, are scalable for high throughput and in case of the Genexus can be translated into clinical diagnostics. By contrast as much as 50 ng was required to amplify repetitive or difficult DNA sequences comprising the gap regions described below.

High quality and high coverage, non-synonymous SNVs and InDels were called and shared variants between the samples were analyzed. Two shared non-synonymous SNVs were identified across all samples; one in ORF11 at position 15,906 and one in ORF61 at position 99,622. Additional SNVs were unique to a particular isolate. These SNVs were present in the samples from both Malawi and the USA, which includes

**Table 3**

The GenBank accession numbers of the 107 publicly available genomes included in generating the amino acid sequence alignment used to build the maximum likelihood phylogenetic trees of both the K1 gene and the LUR of KSHV. Trees were rooted to BC-I (MK733607).

Accession Number	Clinical Presentation	Reference
ERS1615738/UG12 <sup>a</sup>	HIV & Asymptomatic KS	(Sallah et al., 2018a)
ERS1615741/UG13 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615748/UG15 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615752/UG16 <sup>a</sup>	HIV & Asymptomatic KS	(Sallah et al., 2018a)
ERS1615765/UG110 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615766/UG114 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615777/UG118 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615780/UG119 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615783/UG120 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615707/UG126 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615712/UG128 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615725/UG132 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615727/UG133 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615737/UG136 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615761/UG141 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615775/UG145 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615778/UG146 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615786/UG149 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615706/UG155 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615711/UG156 <sup>a</sup>	HIV & Asymptomatic KS	(Sallah et al., 2018a)
ERS1615714/UG157 <sup>a</sup>	HIV & Asymptomatic KS	(Sallah et al., 2018a)
ERS1615719/UG158 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615729/UG162 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615732/UG163 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615736/UG164 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615743/UG166 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615837/UG212 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615860/UG219 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615807/UG222 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
ERS1615813/UG226 <sup>a</sup>	HIV & Asymptomatic KS	(Sallah et al., 2018a)
ERS1615861/UG244 <sup>a</sup>	Asymptomatic KS	(Sallah et al., 2018a)
LC200589	Non-AIDS KS	(Awazawa et al., 2017)
LC200586	Non-AIDS KS	(Awazawa et al., 2017)
LC200587	Non-AIDS KS	(Awazawa et al., 2017)
LC200588	Non-AIDS KS	(Awazawa et al., 2017)
MK733606	KICS	Caro-Vegas et al. (2020)
MK733608	KICS	Caro-Vegas et al. (2020)
KT271453	Classic KS	Olp et al. (2015)
KT271454	Classic KS	Olp et al. (2015)
KT271456	Classic KS	Olp et al. (2015)
KT271455	Classic KS	Olp et al. (2015)
KT271457	Classic KS	Olp et al. (2015)
KT271458	Classic KS	Olp et al. (2015)
KT271459	Classic KS	Olp et al. (2015)
KT271460	Classic KS	Olp et al. (2015)
KT271461	Classic KS	Olp et al. (2015)
KT271462	Classic KS	Olp et al. (2015)
KT271463	Classic KS	Olp et al. (2015)
KT271464	Classic KS	Olp et al. (2015)
KT271465	Classic KS	Olp et al. (2015)
KT271466	Classic KS	Olp et al. (2015)
KT271467	Classic KS	Olp et al. (2015)
KT271468	Classic KS	Olp et al. (2015)
MN419224	PEL and KS	Cornejo Castro et al. (2020)
MN419221	PEL and KS	Cornejo Castro et al. (2020)
MN419226	PEL and KS	Cornejo Castro et al. (2020)
MN419223	PEL and KS	Cornejo Castro et al. (2020)
MN419219	PEL	Cornejo Castro et al. (2020)
MN419225	PEL and KS	Cornejo Castro et al. (2020)
MN419222	PEL/KS/MCD	Cornejo Castro et al. (2020)
MN419227	PEL	Cornejo Castro et al. (2020)
MK876738	MCD	Jary et al. (2020)
MK876737	PEL	Jary et al. (2020)
MK876733	PEL	Jary et al. (2020)
MK876732	KS	Jary et al. (2020)
MK876735	MCD	Jary et al. (2020)
MK876736	PEL	Jary et al. (2020)
MK876734	MCD	Jary et al. (2020)
MT510665	KS	Santiago et al. (2021)
MT510663	KS	Santiago et al. (2021)
MT510664	KS	Santiago et al. (2021)

**Table 3 (continued)**

Accession Number	Clinical Presentation	Reference
MT510669	KS	Santiago et al. (2021)
MT510670	KS	Santiago et al. (2021)
MT510656	KS	Santiago et al. (2021)
MT510658	KS	Santiago et al. (2021)
MT510657	KS	Santiago et al. (2021)
MT510654	KS	Santiago et al. (2021)
MT510655	KS	Santiago et al. (2021)
MT510662	KS	Santiago et al. (2021)
MT510660	KS	Santiago et al. (2021)
MT510659	KS	Santiago et al. (2021)
MT510661	KS	Santiago et al. (2021)
MT510652	KS	Santiago et al. (2021)
MT510653	KS	Santiago et al. (2021)
MT510651	KS	Santiago et al. (2021)
MT510650	KS	Santiago et al. (2021)
MT510648	KS	Santiago et al. (2021)
MT510649	KS	Santiago et al. (2021)
MT510668	KS	Santiago et al. (2021)
MT510666	KS	Santiago et al. (2021)
MT510667	KS	Santiago et al. (2021)
U75698.1/KSU75698	PEL; B cell line	Russo et al. (1996)
JQ619843	MCD and HHV6A	Tamburro et al. (2012)
HQ404500	BAC	Yakushko et al. (2011)
JX228174.1	BAC36deltak15	Bala et al. (2012)
KF588566	HHV8-Brk.219	Kati et al. (2015)
GQ994935.1	JSC1 Clone BAC16	Brulois et al. (2012)
NC_009333	Classic KS	Glenn et al. (1999)
AP017458	PEL cell line	(Osawa et al., 2016)
U93872.2	KS	(Neipel et al., 1997)
MK143395	JSC-1	(De Leo et al., 2019)
MN752405	BAC16orf21_3 GV	Beauclair et al. (2020)
KX189626	BAC16deltaK1	Zhang et al. (2016)
MK497257	BAC16K1 ITAM	Zhang et al. (2016)
KX189629	BAC16K1 REVERTANT	Zhang et al. (2016)
KX189627	BAC16K1 5XSTOP	Zhang et al. (2016)
MK733607	BC-1	Caro-Vegas et al. (2020)

<sup>a</sup> These samples included SRA data that was used to generate consensus sequences using the pipeline described in the methods section.

both the PEL and KS samples and may indicate a sequencing error in the reference sequence. The samples from Malawi had an additional 109 non-synonymous SNVs. These SNVs were incorporated in the reported consensus sequences for each sample and are annotated in the respective GenBank entries.

Of all high confidence non-synonymous SNVs called on these sample 1,324 had a frequency >90% and 91 > 60%. SNVs with <60% minimum frequency were not reported. This result was expected for DNA viruses that replicate via an error-correcting polymerase complex (human polymerase alpha during latency and the viral polymerase ORF9 during lytic replication). Obviously, there must be instances of sequence heterogeneity within individual specimens, otherwise one would not observe evolution; however, in the case of these samples, which represent clonal tumors, such as PEL, and in the case of this virus, which outside of special circumstances (Tamburro et al., 2012) replicates to much lower levels than any other herpesvirus the data set and sequencing depths were too limited to detect these rare events.

### 3.3. Filling in the gaps left by targeted amplification and next generation sequencing

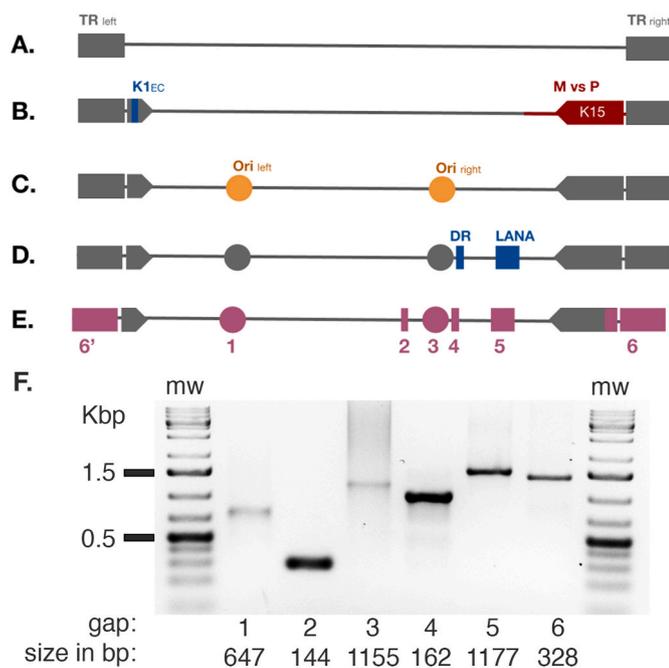
Targeted amplification followed by NGS is very efficient but, like any targeted approach, leaves gaps due to technical difficulties. These gaps are in regions in the genome that cannot be covered by the amplification methods or that are resistant to sequencing-by-synthesis strategies. This problem is not specific to viral sequences. For instance, the human genome, which was declared “completely” sequenced in 1996, currently has ~400,000 contigs, when there should only be 46, one for each chromosome.

Typically, KSHV genomes, even though they are categorized as

complete, only cover the LUR, with the TR either being left out of the sequence entirely or copied from the BCBL-1 or BC-1 strains of KSHV (Lagunoff and Ganem, 1997; Russo et al., 1996). The targeted PCR array used here (AmpliSeq) had six gaps, i.e., regions which could not be covered within the design parameters (melting temperature and size) that were required for multiplexing (Fig. 1). To close the gaps in KSHV in our sequenced assemblies, we designed and tested several primer pairs that target the gap regions left by targeted short-read sequencing. All gap regions could be amplified and sequenced using KSHV-BAC DNA as input (Fig. 1f). The process was less successful for clinical samples. Gap 5, comprised of 1177 bp of the internal repeat domain within the LANA orf, did not always yield unique PCR products and sequence information; all other gap regions did. Like EBV, KSHV encodes two copies of Ori-Lyt, one between ORFs K4.2 and K5, the other between ORF K12 and ORF 71 (Russo et al., 1996; Wang et al., 2004). The two Ori-Lyt sites share an almost identical sequence with a 600 bp GC-rich repeat region represented as 20–30 bp tandem repeat units. The entire region is required for KSHV replication (Wang et al., 2004). The primers termed “gap 1” and “gap 3” cover Ori-L and Ori-R, respectively.

The “gap 1” region (position 24223–24869; gap size = 647 bp, predicted product size = 850bp; average GC content of 88%) was sequenced in BAC DNA. A 692 bp fragment was obtained that covered both DR1 and DR2 repeat regions. PCR products for the gap 1 region in both PEL samples were obtained but we failed to obtain good quality reads by Sanger sequencing.

The “gap 2” (position 113662–113805; gap size = 144bp, PCR product size = 294bp; average GC content of 84%) is a short, GC-rich region within the ORF67 gene. It is specific to the AmpliSeq primer set deployed here. A 294 bp fragment was Sanger sequenced in all samples, with coverage across the entire gap region.



**Fig. 1.** A schematic representation of the AmpliSeq gap regions in relation to genomic architectures. A. depicts the 3' and 5' TR of KSHV. B. The inclusion of the hypervariable K1 and K15 genes that are used to genotype KSHV. C. The left and right origins of replication. D. The DR and LANA. E. An overlay of the AmpliSeq gap regions on the genomic features described in A-D. These include: 1; Left origin of replication, 2; gap in ORF67/ 3; right origin of replication, 4; the DR, 5; LANA and 6; the TRs. Numbers depicted in panel E refer to the gaps in panel F. F. Agarose gel of the gap-spanning amplicons. The gap sizes in base pairs are indicated on the figure. Note that the PCR amplification fragments are larger and correspond to the size in Table 4, since they were optimized to obtain a Tm compatible primer set.

The “gap 3” region (position 117929–119083; gap size = 1155bp, predicted product size = 1345bp; average GC content of 89%) includes the DR5 and DR6 repeats of Ori-R, which follow the Kaposin gene, K12. In BAC DNA, a 581 bp fragment was Sanger sequenced that covered the DR5 repeat region falling short of the entire region, which is 1155 bp. In the PEL patient samples UNC\_RM\_30039264 and UNC\_RM\_30039265, the Sanger read lengths were 699 bp and 791 bp, respectively.

The “gap 4” (position 120334–120496; gap size = 163bp, PCR product size = 1,000bp; average GC content of 96%) includes a GC-rich region that is downstream from the long-interspersed repeats (LIR-1), between K12 and ORF71. In BAC DNA a 963 bp fragment was Sanger sequenced, with coverage across the entire AmpliSeq gap region. Similarly, coverage in the PEL patient samples also covered the AmpliSeq gap region yielding a 873 bp Sanger read for sample UNC\_RM\_30039264 and a 908 bp Sanger read for sample UNC\_RM\_30039265.

The “gap 5” region (position 125146–126322; gap size = 1,177bp, PCR product size = 1,807bp; average GC content of 80%) includes the central region of LANA/ORF73, i.e. includes the large acidic and glutamine-rich internal repeat domain that separates the N- and C-terminus of LANA (Ballestas et al., 1999). The LANA internal repeat domain reportedly has 3 segments containing multiple repeats of the amino acid motif DEED or DEED (amino acids 340–341); a segment of repetitive motifs of QQQEP, QQREP QQQDE (amino acids 760–931), and a segment of multiple repeats of QEQELEE and QELEVEE with amino acids L, V and Q spaced within the region in a leucine zipper-like pattern (Russo et al., 1996). For KSHV-BAC16 DNA a 1372 bp fragment was Sanger sequenced, with coverage along the entire gap region including approximately 80% of the LANA internal repeat domain. The repetitive motifs encoded in the reads from the C-terminal of LANA included in the following order, QQQEP (residues 1325–1743), QQQDE (1744–1884), QEQQEE (1885–1956), QEQELEE (1957–2475), and QELEVEE (2476–2646), with amino acids L, V and Q spaced within the motifs. The length of the LANA IR reportedly varies between different isolates (Rainbow et al., 1997). We tested the primers in various cell types, including BCBL-1, BC1, iSLK219 and HEK293, all of which produced PCR products which were successfully cloned, and Sanger sequenced, covering the entire gap region. Their sequence motifs were in accordance with the KSHV-BAC16 motifs, described above. In the patient samples, a very faint agarose gel band was observed for sample UNC\_RM\_30039264, which only produced poor quality Sanger reads. The PCR amplification was poorly reproducible across clinical samples. This would suggest that biological variation, rather than technical difficulties, were responsible for the result. To test this hypothesis the regions flanking Gap 5 were amplified. Gap 5 lies within the LANA ORF. Therefore, primers were designed to amplify the left and right regions around gap 5. These essentially amplify the entire LANA ORF when a combination of the forward left primer and reverse right primers were used. The regions flanking Gap 5 could be amplified in KSHV-BAC16 DNA, BCBL-1 DNA, and all patient samples, supporting the notion that the GC-rich repeat nature of gap 5 posed technical difficulties, rather than biological variation resulting in internal deletions of LANA. In KSHV-BAC16 DNA all three regions were amplified and Sanger sequenced. In BCBL-1 DNA the left and right regions were amplified, and Sanger sequenced but a product containing the entire LANA ORF could not be obtained. In the PEL patient samples UNC\_RM\_30039264, very faint PCR bands were observed for the left and right regions, but no bands were observed for the complete LANA ORF. In sample UNC\_RM\_30039265, comparably more intense PCR bands were observed for the left and right regions, compared to sample UNC\_RM\_30039264 and with LANA, faint PCR bands were observed. Poor quality Sanger reads were obtained from the patient samples. This could be due to the poor quality of DNA, considering the samples were over 20-year-old FFPE samples and considering that FFPE DNA is subjected to fragmentation and chemical modifications.

The “Gap 6” (position 137767–138094; gap size = 328bp, PCR product size = 1033bp; average GC content of 96%) includes a region

towards the 3'-terminus of the TR of KSHV. The region was completely Sanger sequenced using our primer pairs and in the circular KSHV-BAC16 DNA as input. Reverse primer gap 6 (Table 4) was located at the very 3' end of the linear genome, reverse primer gap 6 within the 5' first ORF K1 to prime backwards across the TR in the circular plasmid form of the virus. In the PEL patient sample, UNC\_RM\_30039264, a 268bp fragment was Sanger sequenced that only just partially covered the AmpliSeq gap region. In the PEL sample UNC\_RM\_30039265, a 588bp fragment was Sanger sequenced with coverage along the entire gap region.

In sum, short-read sequencing rapidly and reliably delivered sequence information for over 90% of the KSHV genome. To obtain 100% complete viral genomes, the GC-rich and repeat regions needed to be individually amplified and sequenced. This was possible using 800 bp Sanger reads on purified, highly concentrated DNA, often requiring an intermediate cloning step. Nevertheless, obtaining long reads remains a challenge for clinical samples, particularly if preserved by fixation with formalin rather than by flash freezing.

Our attempts to sequence KSHV using the Oxford Nanopore technology yielded longer reads but these did not capture these high complexity regions in the viral episome, either. This suggests that their secondary structure inhibits all polymerases that are used in NextGen sequencing under default conditions and requires careful optimization of the reaction for just these regions. This substantiates the usefulness of gap-specific primer pairs.

### 3.4. Comparative analysis of KSHV lineages – K1 and K15

KSHV, like other herpesviruses, has a low mutation frequency and a low recombination frequency. Initial studies distinguished two major lineages based on alleles of the K15 gene on the “right hand” side of the genome. These are the minor “M” allele, represented by the BC-1 PEL cell line, and the predominant “P” or “P1” allele (Poole et al., 1999). On top of this classification, additional subtypes or clades (A, B, C, D, E and F) were defined based on the hypervariable region of K1, which is located on the “left hand” side of the genome. Using previously published K15 representative genotypes of the M (Nicholas et al., 1998) and P (Glenn et al., 1999) allele, including the isolates from Zambia representing the rare N allele (Olp et al., 2015), we characterized the isolates described in this study, with the P allele or subtype, summarized in Table 1 and 2.

**Table 4**

The primer pairs designed to amplify the AmpliSeq gap regions in KSHV-BAC16 DNA, with their validated melting temperatures. The universal M13 primers were tagged at the 5' end of the sequence.

Name (1)	Sequence (3)	Amplicon Size in bp	Tm in °C (2)
Gap 1 fwd	5'-TTCGCCGGGAACGCTATAAAAACG-3'	1052	51.7
Gap 1 rev	5-GGTTATATGCGCGTGCTTC-3'		51.7
Gap 2 fwd	5-CATGACCTTCTCACCAGCGC-3'	294	51.7
Gap 2 rev	5-ATGCCGTACTACTTCACGG-3'		51.7
Gap 3 fwd	5-CAACAGACAAACGAGTGGTGGTATCGC-3'	864	61.7
Gap 3 rev	5'-TACACGTATCGAGGAGCGGT-3'		61.7
Gap 4 fwd	5-CTCTAATCGCTGATTGGTTCCCGC-3'	999	51.7
Gap 4 rev	5-CCACCGATGAGATACCACGCAGC-3'		51.7
Gap 5 fwd	5-GTCCTCGGATGACGACCCG-3'	1825	63.8
Gap 5 rev	5'-GGCTGGCGGAGGATAATGGGG-3'		63.8
Gap 6 fwd	5-GACTAGGTATCCACAGGGCTTAC-3'	1014	51.7
Gap 6 rev	5'-CCCATGCCCGGGGAGGCG-3'		51.7
Gap 6' rev (4)	5'-TAACACCCCTCCGTTTGGTCC-3'	1900	-

1. “fwd” indicated forward and “rev” indicated the reverse primer.
2. Experimentally determined optimal annealing temperature.
3. Based on NC\_009333.
4. Based on GQ994935.

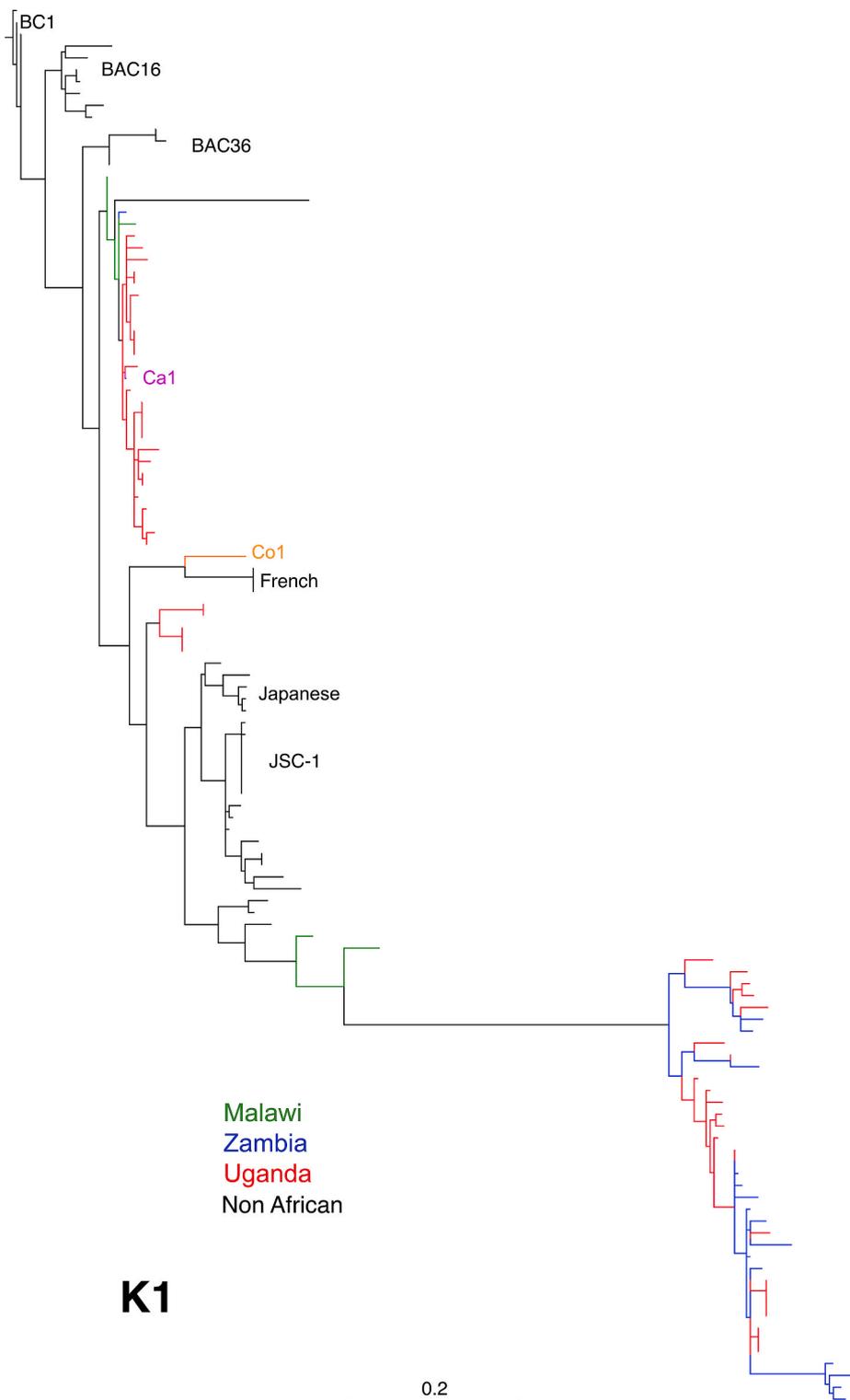
To investigate the variability of the V1 and V2 hypervariable regions of the K1 gene and to generate an updated phylogeny of KSHV, the K1-genes of 131 sequences of previously published genotypes were compared to our samples (Supplementary Table 1). The maximum likelihood phylogenetic tree used to genotype the K1 gene revealed the clustering of the new US PEL samples, namely UNC\_RM\_30035668, UNC\_RM\_30029465 and UNC\_RM\_30004479, along with the US KS samples namely, UNC\_RM\_167 and UNC\_RM\_64, with KSHV strains isolated from the USA in the C3 subtype. The other USA KS isolates in our study, namely UNC\_RM\_162 and USA\_RM\_161 were classified as the A subtypes, consistent with previously isolated strains from the USA (individual genotypes are reported in Table 1 and 2). The isolates from Malawi, including UNC\_RM\_23, UNC\_RM\_98 and UNC\_RM\_52 clustered within the A5 subtype as expected of the African isolates.

The BC-1 rooted maximum likelihood phylogenetic tree built on the K1 gene was constructed using the K1 genes of 107 publicly available genomes isolated from USA, Europe, France, Japan, Uganda, and Zambia, including the BAC sequences referenced in Table 3. The tree topology demonstrated the distinct clustering of the USA isolates described in this study with other isolates from the USA, Japan, and Europe, including the BAC16 and BAC36 sequences (Fig. 2). The African isolates clustered away in two different subtypes from the USA and European, with a few outliers. First, MK876734 (Co1) is an isolate from a Congolese woman with MCD and clustered close with the other French isolates, which were all reportedly genotyped to be a new F subtype of KSHV (Jary et al., 2020). These French isolates clustered close to the African subtype and several genomes isolated from Uganda. The genomes from Uganda, namely MT510663, MT510664, and MT510665, are all genotyped as the C1 subtype and originate from the same patient and include two samples isolated from tumor tissue and one from an oral swab, respectively (Santiago et al., 2021). Clustering within the first African subtype is the genome JQ619843, isolated from a patient in the USA, co-infected with HHV8 and HHV6a, which had not been previously genotyped by its K1 gene (Tamburro et al., 2012). Three of the Malawian genomes described and genotyped in this study to belong to the A subtype, namely UNC\_RM\_23, UNC\_RM\_98 and UNC\_RM\_52 clustered within the first African subtype and close to Ugandan genomes previously genotyped as the A5 subtype. Similarly, the genome MN419227 (Ca1), which is isolated from a PEL patient co-infected with EBV type 2, from Cameroon and belonging to the A5 subtype (Cornejo Castro et al., 2020), clustered with the Ugandan genomes with the A5 subtype. We did not observe the isolated branch represented by the genome ZM004 as reported by Olp et al. (2015). The realignment of the genome ZM004 may be due to the inclusion of an additional 112 genomes which were isolated from various geographical regions including Uganda, Malawi, Zambia, France, and the USA, that were analyzed here as compared to this prior study. All the new isolates from Malawi and the USA, described in this study clustered within the African and USA/European subtypes, respectively.

In sum, the phylogenetic analysis of 117 complete K1 sequences, including, for the first time, many from recently sequenced KSHV isolates from SSA demonstrates the existence of three lineages of KSHV: a US/European lineage representing mostly isolates from the beginning of the AIDS pandemic in the 1990 and two distinct SSA lineages representing more recent isolates of KSHV from areas where KSHV always was endemic and where HIV+ KS represent the second most common cancer in men (after prostate cancer), today.

### 3.5. Comparative analysis of KSHV lineages – LUR

To test the hypothesis that there exist three lineages of KSHV sequences, one representing KSHV diseases from the early AIDS epidemic in Europe and two subtypes representing more recent cases of HIV+ KS in KSHV endemic regions, the phylogenetic analysis was extended to the entire conserved central region of KSHV, the LUR. This region encompasses 67,021 bp of continuous sequence and removes the contribution



**Fig. 2.** Maximum likelihood phylogenetic tree based on 107 publicly available K1 amino acid sequences and the isolates described in this study. The tree is rooted BC-1. The branch colored in orange (MK876734; Co1) includes an isolate from Congo and the branch colored in purple (MN419227; Ca1), is an isolate from Cameroon. Branches colored in black represent genomes from non-African isolates, “French” and “Japanese” isolates are labeled as such. BAC16 and BAC36 refer to the commonly used bacmid clones. Also indicated is the distance scale in the number of substitutions per sites. Maximum likelihood trees were generated using RAxML, with 1000 bootstrap replicates.

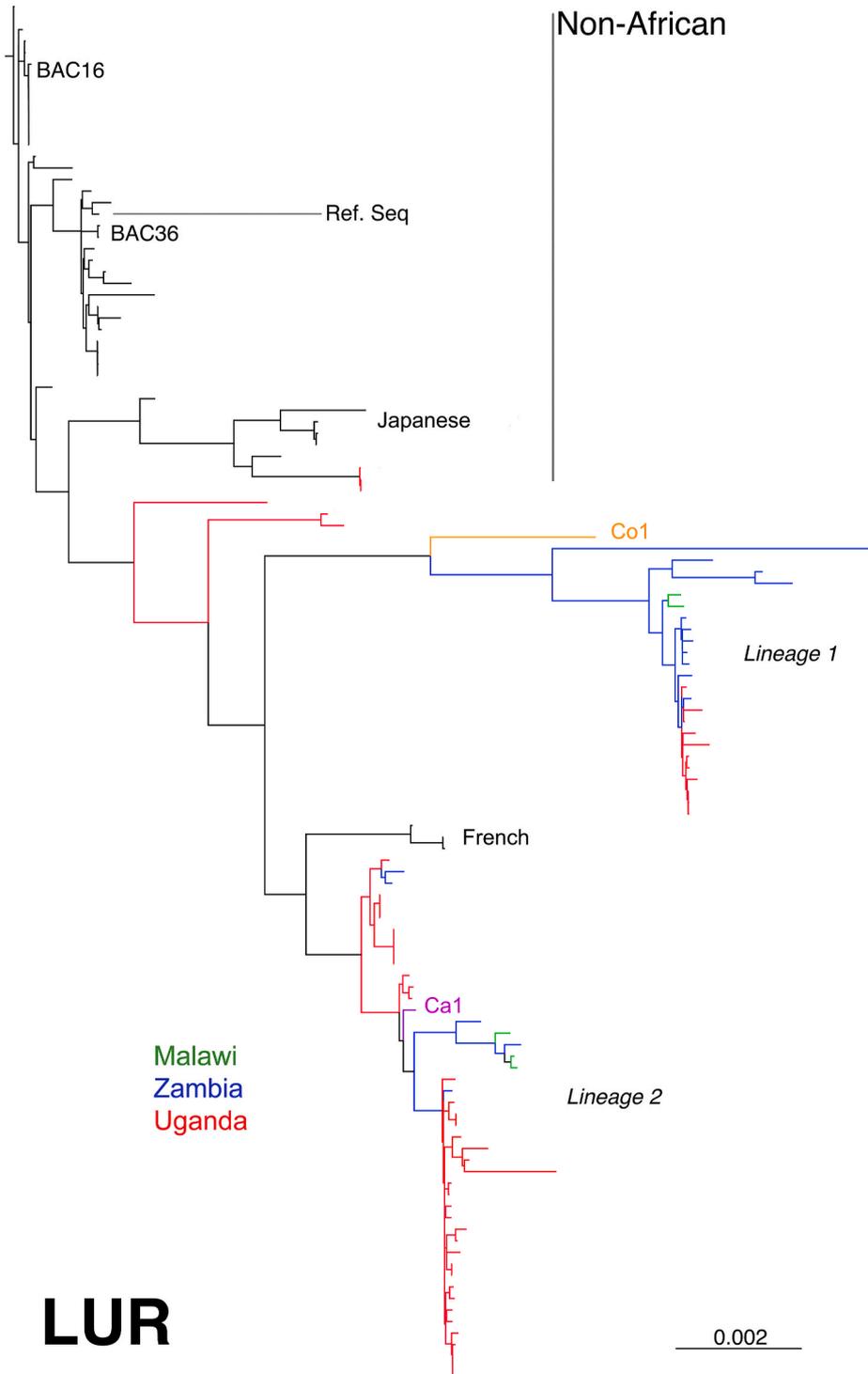
of the hypervariable terminal regions (K1, K15), which are under immune pressure and subject to frequent recombination (Sallah et al., 2018; Zong et al., 1997). Mutational changes in the LUR are more likely attributable to genetic drift and/or selection for replication and pathogenic functions of the virus.

The unrooted phylogenetic ML tree of the central conserved region of KSHV, between the two Ori-Lyt sites, had a striking topology, with all the African isolates clustered in a separate and distinct subtype from the USA and Japanese isolates. Notably all the BAC16 and BAC36 sequences

clustered with the early AIDS US and Japanese isolates, not with any of the more recently sequenced isolates from SSA (Fig. 3). These findings suggest that the LUR of KSHV represents variability, which allows for the geographical clustering of our samples independent of K1-directed immune pressures. Unexpectedly, the five isolates from Uganda that clustered closed to the isolates from Japan are the same genomes that clustered with the European, USA, and BAC sequences based on K1 phylogenetic analysis described above. The LUR of these genomes namely, MT510663, MT510664, MT510665, UG156, and UG157 were

closer related to the Japanese genomes as compared to the Zambian, Malawian and the remaining Ugandan genomes. The Malawian isolates described in this study clustered close to the Zambian and Ugandan isolates, which is consistent with the geographic proximity of the three countries. The French isolates remained clustered within the African subtype, but interestingly the Congolese MCD patient (Co1; MK876734), clustered away from the other French isolates, unlike in the K1 tree. Similarly, to the findings described by Olp et al., the genome ZM004 clustered within the African subtype but had an isolated branch compared to the other Zambian genomes.

The distinct clustering of the African isolates, for instance, in comparison to the KSHV BAC suggested an evolutionary relationship that could be investigated more formally using Bayesian evolutionary analysis as implemented in the BEAST algorithm (Drummond and Rambaut, 2007). The time-dated Bayesian phylogenetic tree confirmed these findings, reproducing the three subtypes; one consisting of sequences isolated from the USA, Europe, Japan, and the BAC sequences. Two subtypes of the African isolates were produced, with the Japanese isolates clustering close to the African subtypes.

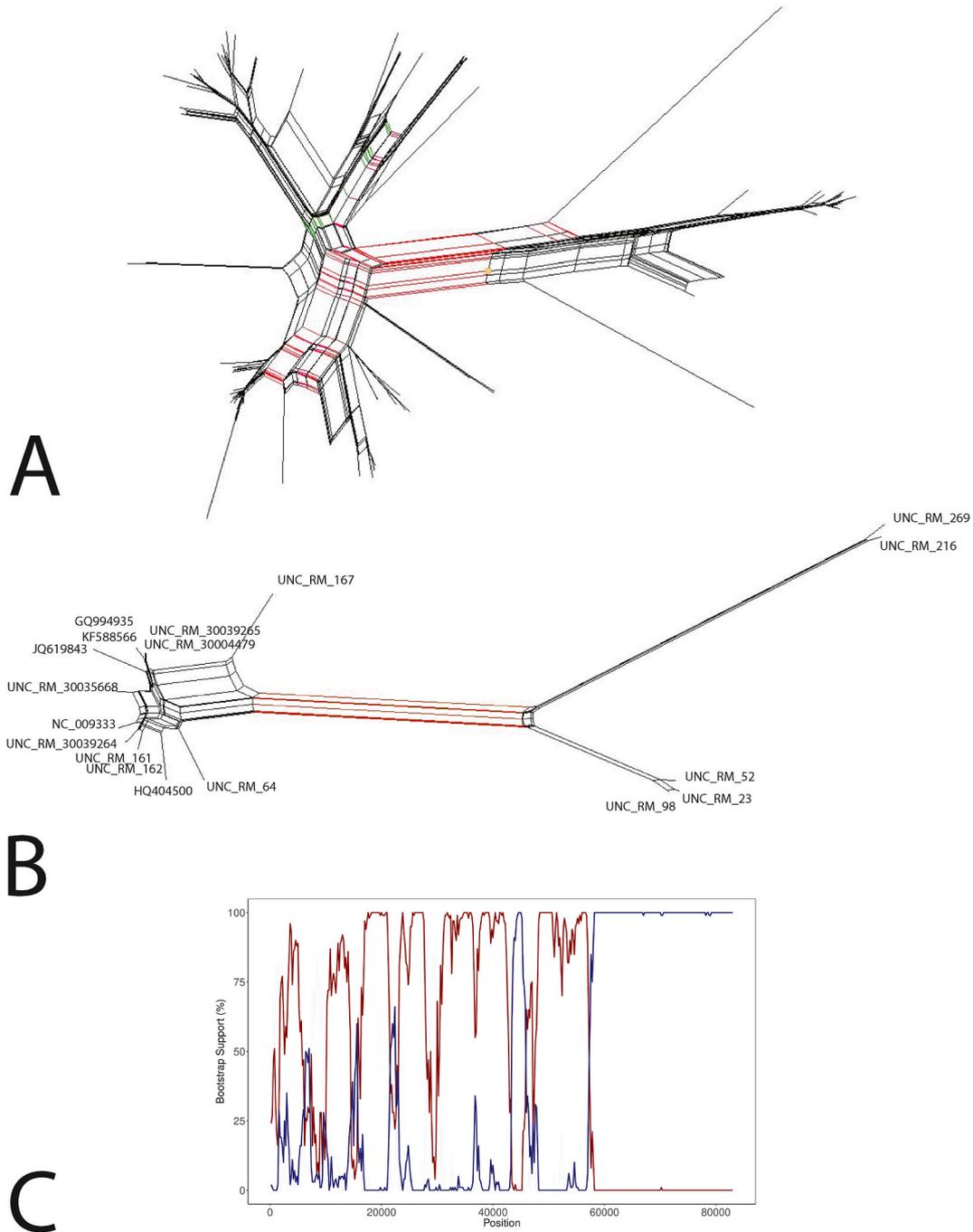


**Fig. 3.** Unrooted maximum likelihood phylogenetic tree of the nucleotide sequences of the central LUR of 107 publicly available genomes aligned to the nucleotide sequence of the LUR of the isolates described in this study. The branch colored in orange (MK876734; Co1), is an isolate from Congo and the branch colored in purple (MN419227; Ca1) is an isolate from Cameroon. Branches colored in black represent genomes from non-African isolates, “French” and “Japanese” isolates are labeled as such. BAC16 and BAC36 refer to the commonly used bacmid clones. “Ref Seq” refers to the current GenBank ICTV reference isolate GK18: NC\_009333. “Lineage 1” and “Lineage 2” refer to the two African branches. Also indicated is the distance scale in the number of substitutions per site. Maximum likelihood trees were generated using RAxML, with 1000 bootstrap replicates.

3.6. Genetic recombination analysis

To formally investigate potential recombination events in our genomes, a SplitsTree network (Kloepper and Huson, 2008) on the LUR was generated. The SplitsTrees network divided the sequences into 3 distinct partitions, similarly to and supporting the phylogenetic clustering of genomes based on their geographical origin (Fig. 4a). The parallel internal lines represent recombination events and/or

convergent evolution. Differentiating sequences can be seen within our dataset, with the splits shown in red representing the genomes from SSA, green splits representing the Japanese genomes and black splits representing the USA/European genomes and BAC sequences. Within the network, a Phi test indicated statistically significant evidence for recombination ( $p < 0.05$ ). Multiple complex relationships within the genomes from SSA were identified, as compared to those from the USA and Europe, which was further investigated by constructing a Splits



**Fig. 4.** A. The Neighbor-Net split network analysis based on the nucleotide alignment of the 117 KSHV-LUR sequences, used to generate the maximum likelihood tree. Parallel lines represent conflicting phylogenetic signals, separating the network into 3 distinct clusters. Splits shown in red represent the African isolates, the splits shown in green represent the isolates from Japan and the black splits representing genomes from USA/Europe. B. The Neighbor-Net split network analysis based on the LUR of the 13 genomes described in this study and 5 reference genomes. The Parallel lines represent conflicting phylogenetic signals, separating the network into 2 distinct clusters. One cluster representing the genomes from the USA, including the reference genome, and the other cluster represents the genomes isolated from Malawi. C. The BootScan analysis depicting evidence of genetic recombination within a genome isolated from Malawi, UNC\_RM\_269. A breakpoint at the approximate position 60,000 bp was identified with the minor parent, UNC\_RM\_52 shown in red and the recombinant, UNC\_RM\_269 shown in blue.

network based on the genomes described in this study and 5 reference genomes (Fig. 4b). This analysis further supported the finding of two African subtypes seen in the LUR phylogenetic analysis, with statistical evidence of recombination ( $p < 0.05$ ). A SimPlot and BootScan analysis were used to determine and visualize the extent of sequence fragmentation and thus, recombination, within the genomes isolated from Malawi, with each genome compared to the reference sequence NC\_009333. The analyses revealed strong statistical support for recombination across four of the five methods used in RDP4, despite the high sequence similarity ( $\geq 98\%$ ). Evidence for intertypic recombination in one of the Malawian genomes, UNC\_RM\_269 was identified, with the minor parent being another genome isolated from Malawi, UNC\_RM\_52 (Fig. 4c).

In sum, this study presents firstly, a new targeted amplification method to sequence KSHV, secondly new tools to obtain 100% complete viral genomes that include the GC-rich viral origins of replication, thirdly a phylogenetic analysis of KSHV evolution based on the largest whole genome data set, yet. This analysis posits the existence of an “African” lineage of KSHV, which represent the viruses that are responsible for 90% of the KS and KSHV-associated lymphoma burden in the world. The recent KSHV isolates largely resemble those used as experimental tools (cell lines and bacmids) that were derived in the early 1990s from US AIDS patients; however, they also present with multiple mutations. It is unclear, which of these are associated with novel phenotypes.

#### 4. Discussion

KS is a leading cancer in PLWH worldwide. In regions where KSHV and HIV are endemic, KS is the most prevalent cancer in both men and women, today (Bray et al., 2018). Despite the uniformly high prevalence of KSHV across SSA, KS disease prevalence shows geographic variation. Endemic KS is mainly observed in Central and Eastern Africa, which may be indicative of genetic variants, observation bias, or regional factors that increase viral pathogenicity and/or transmission (Nalwoga et al., 2020a; Sabourin et al., 2020). This study aimed to investigate recent KSHV evolution based on complete genomic sequences generated here and others that only became available recently.

Analogous to human genetic data supporting a rich genetic diversity within Africa and the migration of modern humans out of Africa, prior analyses of using just the K1 locus of KSHV strongly support specific migration patterns (Nalwoga et al., 2020b). The whole genome based phylogenetic analyses presented here corroborate the general model of herpesvirus-human co-evolution. This model applies to KSHV, prior to the emergence of AIDS in the 1990s (Centers for Disease, 1981), at which point founder effect and sampling bias towards US and European AIDS KS dominate KSHV phylogeny. To date, no KSHV sequences have been reported from transplant KS or from pediatric KS cases.

To facilitate reproducibility and evaluate the robustness of the data presented here it is important to point out some technical details. This study is based on a tiled PCR array (AmpliSeq). Targeted amplification by PCR is independent of any sequencing method. It has greater sensitivity as compared to hybridization-based baits. The unique feature of the AmpliSeq chemistry is that the targeting PCR primers are degraded prior to library construction. This removes a possible sequence bias inherent to other PCR-based methods, where up to 1/5th of the genome sequence may represent input primers, not the target and where primer sequences are purged post facto by bioinformatic methods alone.

This and other reports do not cover “hard-to-sequence” regions of the viral genome, such as the TR or the Ori-Lyt region. Typically, these are presented as gaps in published genomes or imputed from the reference sequence. These “hard-to-sequence” regions nevertheless have functional significance. For instance, high-level replication depends on intact origins of replication. KSHV latent genome maintenance depends on intact TR of a minimal repeat length (Ballestas et al., 1999; Grundhoff and Ganem, 2003; Hu et al., 2002; Hu and Renne, 2005; Renne et al.,

1996a) and the LANA IR region is the predominant target for anti-KSHV antibodies in infected individuals. The GC-rich region downstream from the LIR-1, between K12 and ORF71 encodes the KSHV miRNAs, as well as transcripts antisense to latent lncRNA (Dittmer et al., 1998; Schifano et al., 2017), which is polymorphic in PEL (Sadler et al., 1999). This study reports tools to investigate those “hard-to-sequence” regions and confirms that these non-coding regions are conserved and stably maintained in culture.

The most important result of this study is the realization that KSHV genomes present in KS lesions today, differ from the virus isolates that are used experimentally to understand the biology of KSHV and that originated from US patients in the 1990s prior to the introduction of cART. There are no comparable whole KSHV genome sequences from HIV+ KS biopsies collected in SSA in the 1980. Hence, this data cannot address questions of temporal evolution.

This study demonstrates the existence of two lineages in KS endemic regions. Based on the low mutation rate of herpesviruses and extensive co-evolution with the human host, one would assume that essential viral functions, i.e., potential therapy targets, nevertheless have been conserved. Conversely, an expanded phylogenetic investigation, may uncover new biological determinants, akin to BRLF1 promoter polymorphism in EBV (Bhende et al., 2004). At this point, none of the KSHV sequence variation reported here has been linked to functional differences, clinically or in culture.

The K1 tree (Fig. 2) like the whole genome tree (Fig. 3) were congruent in identifying two distinct African lineages. Whereas K1-based phylogenies are able to resolve subsets within the major lineages, the sparseness of polymorphisms outside of K1 did not have the same resolving power.

A limiting factor to the recombination analysis in this study, is the overall small sample size of available KSHV genomes. This study included the largest number of KSHV genomes used to investigate the evolutionary relationships of KSHV. Tree topology will inevitably change with the inclusion of additional full KSHV genomes. Whilst evidence of recombination has been shown, it is unknown whether these genomes existed as parental recombinant strains prior to infecting a particular individual or if recombination was a consequence of ongoing co-infection at the time of sampling. Since the KSHV genome is highly conserved and this approach uses short read sequencing technology, differentiating recombination between similar strains from PCR errors with high statistical confidence posed a challenge. Within the constraints of our experimental approach, we could only detect one recombination event with sufficient significance (Fig. 4C). One the one hand, this was expected, since our samples were from predominantly latent KS lesions and in the case of PEL from clonal tumor cells. One would expect these samples to be dominated by a single viral clone, as opposed to saliva samples, which contain populations of cells that actively replicate virus and in which recombination is ongoing as part of herpesvirus packaging. One the other hand it is quite possible that higher sequencing depth using Unique Molecular Index (UMI) technology would be more sensitive to detect co-infections and inter-strain recombination.

By comparison to EBV and other human herpesviruses, very few KSHV genome sequences are available. This represents a barrier to put experimental insights into a global perspective. It hampers translational research, vaccine, and therapy development. The sampling bias that this study uncovered puts those who experience the highest burden of disease at a disadvantage. We and others continue to isolate and sequence KSHV from patient samples obtained from various geographical regions including USA, Europe, Japan, Zambia, Uganda, and Malawi (Caro-Vegas et al., 2020; Dittmer and Damania, 2019; El-Mallawany et al., 2019a; Olp et al., 2015; Tamburro et al., 2012). This study would not have been possible without the prior work conducted by others in the field and their willingness to publicly share genomic sequences.

## Role of the funding source

This work was supported by US Public Health Service grant CA239583 to DPD, 5UM1CA181255 to the AIDS Cancer Specimen Resource (ACSR), and an AIDS and Malignancy Consortium (AMC) fellowship to RM (2-UM1-CA121947). The funders had no influence on study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

## Availability of data and materials

All genomes and/or datasets analyzed and generated in this study are available in the GenBank repository (<https://www.ncbi.nlm.nih.gov/genbank/>), under the accession numbers listed in Tables 1 and 2.

## CRedit authorship contribution statement

**Razia Moorad:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. **Angelica Juarez:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology. **Justin T. Landis:** Validation, Software, Resources, Data curation. **Linda J. Pluta:** Investigation. **Megan Perkins:** Investigation. **Avery Cheves:** Investigation. **Dirk P. Dittmer:** Conceptualization, Validation, Formal analysis, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We thank the members of the Dittmer laboratory for their support, discussions, and critical reading.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.virol.2022.01.005>.

## References

- Arias, C., Weisburd, B., Stern-Ginossar, N., Mercier, A., Madrid, A.S., Bellare, P., Holdorf, M., Weissman, J.S., Ganem, D., 2014. KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog.* 10, e1003847.
- Awazawa, R., Daisuke, U., Harutaka, K., Awazawa, T., Miyagi, T., Hayashi, K., Matori, S., Uezato, H., Takahashi, K., 2017. 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.* 216 (7), 850–858.
- Bai, Z., Huang, Y., Li, W., Zhu, Y., Jung, J.U., Lu, C., Gao, S.J., 2014. Genomewide mapping and screening of Kaposi's sarcoma-associated herpesvirus (KSHV) 3' untranslated regions identify bicistronic and polycistronic viral transcripts as frequent targets of KSHV microRNAs. *J. Virol.* 88, 377–392.
- Bala, K., Bosco, R., Gramolelli, S., Haas, D.A., Kati, S., Pietrek, M., Havemeier, A., Yakushko, Y., Singh, V.V., Dittrich-Breiholz, O., Kracht, M., Schulz, T.F., 2012. Kaposi's sarcoma herpesvirus K15 protein contributes to virus-induced angiogenesis by recruiting PLCgamma and activating NFAT1-dependent RCAN1 expression. *PLoS Pathog.* 8, e1002927.
- Ballestas, M.E., Chatis, P.A., Kaye, K.M., 1999. Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen. *Science* 284, 641–644.
- Beauchair, G., Naimo, E., Dubich, T., Ruckert, J., Koch, S., Dhingra, A., Wirth, D., Schulz, T.F., 2020. Targeting Kaposi's Sarcoma-associated herpesvirus ORF21 tyrosine Kinase and viral lytic reactivation by tyrosine Kinase inhibitors approved for clinical use. *J. Virol.* 94, e01791-01719.
- Bhende, P.M., Seaman, W.T., Delecluse, H.J., Kenney, S.C., 2004. The EBV lytic switch protein, Z, preferentially binds to and activates the methylated viral genome. *Nat. Genet.* 36, 1099–1104.
- Bigi, R., Landis, J.T., An, H., Caro-Vegas, C., Raab-Traub, N., Dittmer, D.P., 2018. Epstein-Barr virus enhances genome maintenance of Kaposi sarcoma-associated herpesvirus. *Proc. Natl. Acad. Sci. U. S. A.* 115, E11379–E11387.
- Bray, F., Ferlay, J., Soerjomataram, L., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca - Cancer J. Clin.* 68, 394–424.
- Brulois, K.F., Chang, H., Lee, A.S., Ensser, A., Wong, L.Y., Toth, Z., Lee, S.H., Lee, H.R., Myoung, J., Ganem, D., Oh, T.K., Kim, J.F., Gao, S.J., Jung, J.U., 2012. Construction and manipulation of a new Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome clone. *J. Virol.* 86, 9708–9720.
- Budt, M., Hristozova, T., Hille, G., Berger, K., Brune, W., 2011. Construction of a lytically replicating Kaposi's sarcoma-associated herpesvirus. *J. Virol.* 85, 10415–10420.
- Cannon, J.S., Ciufu, D., Hawkins, A.L., Griffin, C.A., Borowitz, M.J., Hayward, G.S., Ambinder, R.F., 2000. A new primary effusion lymphoma-derived cell line yields a highly infectious Kaposi's sarcoma herpesvirus-containing supernatant. *J. Virol.* 74, 10187–10193.
- Caro-Vegas, C., Sellers, S., Host, K.M., Seltzer, J., Landis, J., Fischer, W.A., Damania, B., Dittmer, D.P., 2020. Runaway Kaposi Sarcoma-associated herpesvirus replication correlates with systemic IL-10 levels. *Virology* 539, 18–25.
- Centers for Disease, C., 1981. Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men—New York City and California. *MMWR Morb. Mortal. Wkly. Rep.* 30, 305–308.
- Cesarman, E., Moore, P.S., Rao, P.H., Inghirami, G., Knowles, D.M., Chang, Y., 1995. 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* 86, 2708–2714.
- Chang, H., Wachtman, L.M., Pearson, C.B., Lee, J.S., Lee, H.R., Lee, S.H., Vieira, J., Mansfield, K.G., Jung, J.U., 2009. Non-human primate model of Kaposi's sarcoma-associated herpesvirus infection. *PLoS Pathog.* 5, e1000606.
- Cook, P.M., Whitby, D., Calabro, M.L., Luppi, M., Kakoola, D.N., Hjalgrim, H., Ariyoshi, K., Ensoli, B., Davison, A.J., Schulz, T.F., 1999. Variability and evolution of Kaposi's sarcoma-associated herpesvirus in Europe and Africa. *International Collaborative Group AIDS* 13, 1165–1176.
- Cornejo Castro, E.M., Marshall, V., Lack, J., Lurain, K., Immonen, T., Labo, N., Fisher, N. C., Ramaswami, R., Polizzotto, M.N., Keele, B.F., Yarchoan, R., Uldrick, T.S., Whitby, D., 2020. Dual infection and recombination of Kaposi sarcoma herpesvirus revealed by whole-genome sequence analysis of effusion samples. *Virus Evol.* 6, veaa047.
- De Leo, A., Deng, Z., Vladimirova, O., Chen, H.S., Dheekollu, J., Calderon, A., Myers, K. A., Hayden, J., Keeney, F., Kaufer, B.B., Yuan, Y., Robertson, E., Lieberman, P.M., 2019. LANA oligomeric architecture is essential for KSHV nuclear body formation and viral genome maintenance during latency. *PLoS Pathog.* 15 (1).
- Decker, L.L., Shankar, P., Khan, G., Freeman, R.B., Dezube, B.J., Lieberman, J., Thorley-Lawson, D.A., 1996. The Kaposi sarcoma-associated herpesvirus (KSHV) is present as an intact latent genome in KS tissue but replicates in the peripheral blood mononuclear cells of KS patients. *J. Exp. Med.* 184, 283–288.
- Dittmer, D., Lagunoff, M., Renne, R., Staskus, K., Haase, A., Ganem, D., 1998. A cluster of latently expressed genes in Kaposi's sarcoma-associated herpesvirus. *J. Virol.* 72, 8309–8315.
- Dittmer, D., Stoddart, C., Renne, R., Linguist-Stepps, V., Moreno, M.E., Bare, C., McCune, J.M., Ganem, D., 1999. Experimental transmission of Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) to SCID-hu Thy/Liv mice. *J. Exp. Med.* 190, 1857–1868.
- Dittmer, D.P., 2003. Transcription profile of Kaposi's sarcoma-associated herpesvirus in primary Kaposi's sarcoma lesions as determined by real-time PCR arrays. *Cancer Res.* 63, 2010–2015.
- Dittmer, D.P., Damania, B., 2016. Kaposi sarcoma-associated herpesvirus: immunobiology, oncogenesis, and therapy. *J. Clin. Invest.* 126, 3165–3175.
- Dittmer, D.P., Damania, B., 2019. Kaposi's Sarcoma-associated herpesvirus (KSHV)-Associated disease in the AIDS patient: an update. *Cancer Treat Res.* 177, 63–80.
- Dittmer, D.P., Tamburro, K., Chen, H., Lee, A., Sanders, M.K., Wade, T.A., Napravnik, S., Webster-Cyriaque, J., Ghannoum, M., Shiboski, C.H., Aberg, J.A., 2017. Oral shedding of herpesviruses in HIV-infected patients with varying degrees of immune status. *AIDS* 31, 2077–2084.
- Dresang, L.R., Teuton, J.R., Feng, H., Jacobs, J.M., Camp, D.G., Purvine, S.O., Gritsenko, M.A., Li, Z., Smith, R.D., Sugden, B., Moore, P.S., Chang, Y., 2011. Coupled transcriptome and proteome analysis of human lymphotropic tumor viruses: insights on the detection and discovery of viral genes. *BMC Genom.* 12, 625.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973.
- El-Mallawany, N.K., Kamiyango, W., Villiera, J., Peckham-Gregory, E.C., Scheurer, M.E., McAtee, C.L., Allen, C.E., Kovarik, C.L., Frank, D., Eason, A.B., Caro-Vegas, C., Chiao, E.Y., Schutze, G.E., Ozuah, N.W., Mehta, P.S., Kazembe, P.N., Dittmer, D.P., 2019a. Kaposi Sarcoma herpesvirus inflammatory cytokine Syndrome-like clinical presentation in human immunodeficiency virus-infected children in Malawi. *Clin. Infect. Dis.* 69, 2022–2025.
- El-Mallawany, N.K., Mehta, P.S., Kamiyango, W., Villiera, J., Peckham-Gregory, E.C., Kampani, C., Krysiak, R., Sanders, M.K., Caro-Vegas, C., Eason, A.B., Ahmed, S., Schutze, G.E., Martin, S.C., Kazembe, P.N., Scheurer, M.E., Dittmer, D.P., 2019b. KSHV viral load and Interleukin-6 in HIV-associated pediatric Kaposi sarcoma-

- Exploring the role of lytic activation in driving the unique clinical features seen in endemic regions. *Int. J. Cancer* 144, 110–116.
- El-Mallaway, N.K., Villiera, J., Kamiyango, W., Peckham-Gregory, E.C., Scheurer, M.E., Allen, C.E., McAtee, C.L., Legarreta, A., Dittmer, D.P., Kovarik, C.L., Chiao, E.Y., Martin, S.C., Ozuah, N.W., Mehta, P.S., Kazembe, P.N., 2018. Endemic Kaposi sarcoma in HIV-negative children and adolescents: an evaluation of overlapping and distinct clinical features in comparison with HIV-related disease. *Infect. Agents Cancer* 13, 33.
- Gao, S.J., Zhang, Y.J., Deng, J.H., Rabkin, C.S., Flore, O., Jenson, H.B., 1999. Molecular polymorphism of Kaposi's sarcoma-associated herpesvirus (Human herpesvirus 8) latent nuclear antigen: evidence for a large repertoire of viral genotypes and dual infection with different viral genotypes. *J. Infect. Dis.* 180, 1466–1476.
- Glenn, M., Rainbow, L., Aurade, F., Davison, A., Schulz, T.F., 1999. 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.* 73, 6953–6963.
- Godfrey, A., Anderson, J., Papanastasiou, A., Takeuchi, Y., Boshoff, C., 2005. Inhibiting primary effusion lymphoma by lentiviral vectors encoding short hairpin RNA. *Blood* 105, 2510–2518.
- Grundhoff, A., Ganem, D., 2003. The latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus permits replication of terminal repeat-containing plasmids. *J. Virol.* 77, 2779–2783.
- Hayward, G.S., 1999. KSHV strains: the origins and global spread of the virus. *Semin. Cancer Biol.* 9, 187–199.
- Hosseiniour, M.C., Sweet, K.M., Xiong, J., Namarika, D., Mwafongo, A., Nyirenda, M., Chiwoko, L., Kamwendo, D., Hoffman, I., Lee, J., Phiri, S., Vahrson, W., Damania, B., Dittmer, D.P., 2014. Viral profiling identifies multiple subtypes of Kaposi's sarcoma. *mBio* 5, e01633-01614.
- Hu, J., Garber, A.C., Renne, R., 2002. The latency-associated nuclear antigen of Kaposi's sarcoma-associated herpesvirus supports latent DNA replication in dividing cells. *J. Virol.* 76, 11677–11687.
- Hu, J., Renne, R., 2005. Characterization of the minimal replicator of Kaposi's sarcoma-associated herpesvirus latent origin. *J. Virol.* 79, 2637–2642.
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267.
- Jain, V., Plaisance-Bonstaff, K., Sangani, R., Lanier, C., Dolce, A., Hu, J., Brulois, K., Haecker, I., Turner, P., Renne, R., Krueger, B., 2016. A toolbox for herpesvirus miRNA research: construction of a complete set of KSHV miRNA deletion mutants. *Viruses* 8, 54.
- Jary, A., Leducq, V., Desire, N., Petit, H., Palich, R., Joly, V., Canestri, A., Gothland, A., Lambert-Niclot, S., Surgers, L., Amiel, C., Descamps, D., Spano, J.P., Katlama, C., Calvez, V., Marcelin, A.G., 2020. New Kaposi's sarcoma-associated herpesvirus variant in men who have sex with men associated with severe pathologies. *J. Infect. Dis.* 222, 1320–1328.
- Jones, D., Ballestas, M.E., Kaye, K.M., Gulizia, J.M., Winters, G.L., Fletcher, J., Scadden, D.T., Aster, J.C., 1998. Primary-effusion lymphoma and Kaposi's sarcoma in a cardiac-transplant recipient. *N. Engl. J. Med.* 339, 444–449.
- Jones, T., Ye, F., Bedolla, R., Huang, Y., Meng, J., Qian, L., Pan, H., Zhou, F., Moody, R., Wagner, B., Arar, M., Gao, S.J., 2012. Direct and efficient cellular transformation of primary rat mesenchymal precursor cells by KSHV. *J. Clin. Invest.* 122, 1076–1081.
- Kajumbula, H., Wallace, R.G., Zong, J.C., Hokello, J., Sussman, N., Simms, S., Rockwell, R.F., Pozos, R., Hayward, G.S., Boto, W., 2006. Ugandan Kaposi's sarcoma-associated herpesvirus phylogeny: evidence for cross-ethnic transmission of viral subtypes. *Intervirology* 49, 133–143.
- Kati, S., Hage, E., Mynarek, M., Ganzemueller, T., Indenbirken, D., Grundhoff, A., Schulz, T.F., 2015. Generation of high-titre virus stocks using BrK219, a B-cell line infected stably with recombinant Kaposi's sarcoma-associated herpesvirus. *J. Virol. Methods* 217, 79–86.
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780.
- Kedes, D.H., Lagunoff, M., Renne, R., Ganem, D., 1997. Identification of the gene encoding the major latency-associated nuclear antigen of the Kaposi's sarcoma-associated herpesvirus. *J. Clin. Invest.* 100, 2606–2610.
- Klein, U., Gloghini, A., Gaidano, G., Chadburn, A., Cesarman, E., Dalla-Favera, R., Carbone, A., 2003. Gene expression profile analysis of AIDS-related primary effusion lymphoma (PEL) suggests a plasmablastic derivation and identifies PEL-specific transcripts. *Blood* 101, 4115–4121.
- Klopper, T.H., Huson, D.H., 2008. Drawing explicit phylogenetic networks and their integration into SplitsTree. *BMC Evol. Biol.* 8, 22.
- Krown, S.E., Lee, J.Y., Dittmer, D.P., Consortium, A.M., 2008. More on HIV-associated Kaposi's sarcoma. *N. Engl. J. Med.* 358, 535–536 author reply 536.
- Lagunoff, M., Ganem, D., 1997. The structure and coding organization of the genomic termini of Kaposi's sarcoma-associated herpesvirus. *Virology* 236, 147–154.
- Majerciak, V., Ni, T., Yang, W., Meng, B., Zhu, J., Zheng, Z.M., 2013. A viral genome landscape of RNA polyadenylation from KSHV latent to lytic infection. *PLoS Pathog.* 9, e1003749.
- Maurer, T., Ponte, M., Leslie, K., 2007. HIV-associated Kaposi's sarcoma with a high CD4 count and a low viral load. *N. Engl. J. Med.* 357, 1352–1353.
- McHugh, D., Caduff, N., Barros, M.H.M., Ramer, P.C., Raykova, A., Murer, A., Landtwing, V., Quast, I., Styles, C.T., Spohn, M., Fowotade, A., Delecluse, H.J., Papoudou-Bai, A., Lee, Y.M., Kim, J.M., Middeldorp, J., Schulz, T.F., Cesarman, E., Zbinden, A., Capaul, R., White, R.E., Allday, M.J., Niedobitek, G., Blackbourn, D.J., Grundhoff, A., Munz, C., 2017. Persistent KSHV infection increases EBV-associated tumor formation in vivo via enhanced EBV lytic gene expression. *Cell Host Microbe* 22, 61–73 e67.
- Meng, Y.X., Spira, T.J., Bhat, G.J., Birch, C.J., Druce, J.D., Edlin, B.R., Edwards, R., Gunthel, C., Newton, R., Stamey, F.R., Wood, C., Pellett, P.E., 1999. Individuals from North America, Australasia, and Africa are infected with four different genotypes of human herpesvirus 8. *Virology* 261, 106–119.
- Moore, P.S., Gao, S.J., Dominguez, G., Cesarman, E., Lungu, O., Knowles, D.M., Garber, R., Pellett, P.E., McGeoch, D.J., Chang, Y., 1996. Primary characterization of a herpesvirus agent associated with Kaposi's sarcoma. *J. Virol.* 70, 549–558.
- Nalwoga, A., Nakibuule, M., Marshall, V., Miley, W., Labo, N., Cose, S., Whitby, D., Newton, R., 2020a. Risk factors for Kaposi's Sarcoma-associated herpesvirus DNA in blood and in Saliva in Rural Uganda. *Clin. Infect. Dis.* 71, 1055–1062.
- Nalwoga, A., Webb, E.L., Muserere, C., Chihota, B., Miley, W., Labo, N., Elliott, A., Cose, S., Whitby, D., Newton, R., 2020b. Variation in KSHV prevalence between geographically proximate locations in Uganda. *Infect. Agents Cancer* 15, 49.
- Neipel, F., Albrecht, J.C., Fleckenstein, B., 1997. Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? *J. Virol.* 71 (6), 4187–4192.
- Nicholas, J., Zong, J.C., Alcendor, D.J., Ciuffo, D.M., Poole, L.J., Sarisky, R.T., Chiou, C.J., Zhang, X., Wan, X., Guo, H.G., Reitz, M.S., Hayward, G.S., 1998. Novel organizational features, captured cellular genes, and strain variability within the genome of KSHV/HHV8. *J. Natl. Cancer Inst. Monogr.* 79–88.
- Nyirenda, M., Ngongondo, M., Kang, M., Umbleja, T., Krown, S.E., Godfrey, C., Samaneka, W., Mngqibisa, R., Hoagland, B., Mwelase, N., Caruso, S., Martinez-Maza, O., Dittmer, D.P., Borok, M., Hosseinipour, M.C., Campbell, T.B., team, A.A.-., 2020. Early progression and immune reconstitution inflammatory Syndrome during treatment of mild-to-moderate Kaposi Sarcoma in Sub-Saharan Africa and South America: incidence, long-term outcomes, and effects of early chemotherapy. *J. Acquir. Immune Defic. Syndr.* 84, 422–429.
- Olp, L.N., Jeanniard, A., Marimo, C., West, J.T., Wood, C., 2015. Whole-genome Sequencing of Kaposi's Sarcoma-associated herpesvirus from Zambian Kaposi's Sarcoma biopsy specimens reveals unique viral diversity. *J. Virol.* 89, 12299–12308.
- Osawa, M., Mine, S., Ota, S., Kato, K., Sekizuka, T., Kuroda, M., Kataoka, M., Fukumoto, H., Sato, Y., Kanno, T., Hasegawa, H., Ueda, K., Fukayama, M., Maeda, T., Kanoh, S., Kawana, A., Fujikura, Y., Katano, H., 2016. Establishing and characterizing a new primary effusion lymphoma cell line harboring Kaposi's sarcoma-associated herpesvirus. *Infect. Agent Cancer* 11 (37).
- Poole, L.J., Zong, J.C., Ciuffo, D.M., Alcendor, D.J., Cannon, J.S., Ambinder, R., Orenstein, J.M., Reitz, M.S., Hayward, G.S., 1999. 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.* 73, 6646–6660.
- Rainbow, L., Platt, G.M., Simpson, G.R., Sarid, R., Gao, S.J., Stoiber, H., Herrington, C.S., Moore, P.S., Schulz, T.F., 1997. The 222- to 234-kilodalton latent nuclear protein (LNA) of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) is encoded by orf73 and is a component of the latency-associated nuclear antigen. *J. Virol.* 71, 5915–5921.
- Renne, R., Blackbourn, D., Whitby, D., Levy, J., Ganem, D., 1998. Limited transmission of Kaposi's sarcoma-associated herpesvirus in cultured cells. *J. Virol.* 72, 5182–5188.
- Renne, R., Lagunoff, M., Zhong, W., Ganem, D., 1996a. The size and conformation of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in infected cells and virions. *J. Virol.* 70, 8151–8154.
- Renne, R., Zhong, W., Herndier, B., McGrath, M., Abbey, N., Kedes, D., Ganem, D., 1996b. Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. *Nat. Med.* 2, 342–346.
- Royston, L., Isnard, S., Calmy, A., Routy, J.P., 2021. Kaposi sarcoma in antiretroviral therapy-treated people with HIV: a wake-up call for research on human herpesvirus-8. *AIDS* 35, 1695–1699.
- Russo, J.J., Bohenzky, R.A., Chien, M.C., Chen, J., Yan, M., Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y., Moore, P.S., 1996. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc. Natl. Acad. Sci. U. S. A.* 93, 14862–14867.
- Sabourin, K.R., Daud, I., Ogolla, S., Labo, N., Miley, W., Lamb, M., Newton, R., Whitby, D., Rochford, R., 2020. Malaria is associated with Kaposi sarcoma-associated herpesvirus (KSHV) seroconversion in a cohort of western Kenyan children. *J. Infect. Dis.* 224, 303–311.
- Sadler, R., Wu, L., Forghani, B., Renne, R., Zhong, W., Herndier, B., Ganem, D., 1999. A complex translational program generates multiple novel proteins from the latently expressed kaposin (K12) locus of Kaposi's sarcoma-associated herpesvirus. *J. Virol.* 73, 5722–5730.
- Sallah, N., Palser, A.L., Watson, S.J., Labo, N., Asiki, G., Marshall, V., Newton, R., Whitby, D., Kellam, P., Barroso, I., 2018. Genome-wide sequence analysis of Kaposi Sarcoma-associated herpesvirus shows diversification driven by recombination. *J. Infect. Dis.* 218, 1700–1710.
- Santiago, J.C., Goldman, J.D., Zhao, H., Pankow, A.P., Okuku, F., Schmitt, M.W., Chen, L.H., Hill, C.A., Casper, C., Phipps, W.T., Mullins, J.I., 2021. Intra-host changes in Kaposi sarcoma-associated herpesvirus genomes in Ugandan adults with Kaposi sarcoma. *PLoS Pathog.* 17, e1008594.
- Sarid, R., Flore, O., Bohenzky, R.A., Chang, Y., Moore, P.S., 1998. Transcription mapping of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) genome in a body cavity-based lymphoma cell line (BC-1). *J. Virol.* 72, 1005–1012.
- Schifano, J.M., Corcoran, K., Kelkar, H., Dittmer, D.P., 2017. Expression of the antisense-to-latency transcript long noncoding RNA in Kaposi's Sarcoma-associated herpesvirus. *J. Virol.* 91, e01698-01616.
- Schneider, J.W., Dittmer, D.P., 2017. Diagnosis and treatment of Kaposi Sarcoma. *Am. J. Clin. Dermatol.* 18, 529–539.
- Shiboski, C.H., Chen, H., Secours, R., Lee, A., Webster-Cyriaque, J., Ghannoum, M., Evans, S., Bernard, D., Reznik, D., Dittmer, D.P., Hosey, L., Severe, P., Aberg, J.A.,

- Oral Hiv/Aids Research Alliance, S.o.t.A.C.T.G., 2015. High accuracy of common HIV-related oral disease diagnoses by non-oral health specialists in the AIDS clinical trial group. *PLoS One* 10, e0131001.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Tamburro, K.M., Yang, D., Poisson, J., Fedoriw, Y., Roy, D., Lucas, A., Sin, S.H., Malouf, N., Moylan, V., Damania, B., Moll, S., van der Horst, C., Dittmer, D.P., 2012. 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* 433, 220–225.
- Wang, L.X., Kang, G., Kumar, P., Lu, W., Li, Y., Zhou, Y., Li, Q., Wood, C., 2014. Humanized-BLT mouse model of Kaposi's sarcoma-associated herpesvirus infection. *Proc. Natl. Acad. Sci. U. S. A.* 111, 3146–3151.
- Wang, Y., Li, H., Chan, M.Y., Zhu, F.X., Lukac, D.M., Yuan, Y., 2004. Kaposi's sarcoma-associated herpesvirus ori-Lyt-dependent DNA replication: cis-acting requirements for replication and ori-Lyt-associated RNA transcription. *J. Virol.* 78, 8615–8629.
- Xu, M., Yao, Y., Chen, H., Zhang, S., Cao, S.M., Zhang, Z., Luo, B., Liu, Z., Li, Z., Xiang, T., He, G., Feng, Q.S., Chen, L.Z., Guo, X., Jia, W.H., Chen, M.Y., Zhang, X., Xie, S.H., Peng, R., Chang, E.T., Pedergrana, V., Feng, L., Bei, J.X., Xu, R.H., Zeng, M.S., Ye, W., Adami, H.O., Lin, X., Zhai, W., Zeng, Y.X., Liu, J., 2019. Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma. *Nat. Genet.* 51, 1131–1136.
- Yakushko, Y., Hackmann, C., Gunther, T., Ruckert, J., Henke, M., Koste, L., Alkharsah, K., Bohne, J., Grundhoff, A., Schulz, T.F., Henke-Gendo, C., 2011. Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome contains a duplication of a long unique-region fragment within the terminal repeat region. *J. Virol.* 85, 4612–4617.
- Zhang, Z., Chen, W., Sanders, M.K., Brulois, K.F., Dittmer, D.P., Damania, B., 2016. The K1 protein of Kaposi's Sarcoma-associated herpesvirus augments viral lytic replication. *J. Virol.* 90, 7657–7666.
- Zhou, F.C., Zhang, Y.J., Deng, J.H., Wang, X.P., Pan, H.Y., Hettler, E., Gao, S.J., 2002. Efficient infection by a recombinant Kaposi's sarcoma-associated herpesvirus cloned in a bacterial artificial chromosome: application for genetic analysis. *J. Virol.* 76, 6185–6196.
- Zong, J.C., Metroka, C., Reitz, M.S., Nicholas, J., Hayward, G.S., 1997. 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.* 71, 2505–2511.