HHS Public Access

Author manuscript

Int J Cancer. Author manuscript; available in PMC 2023 April 01.

Published in final edited form as:

Int J Cancer. 2022 October 01; 151(7): 1127–1141. doi:10.1002/ijc.34136.

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

Vickie A. Marshall[1], **Nicholas C. Fisher**[1], **Charles A. Goodman**[2], **Elena M. Cornejo Castro**[1], **Isabella Liu**[1], **Sirish Khanal**[2], **Benjamin M. Holdridge**[1], **Abigail L. Thorpe**[2], **Nazzarena Labo**[1], **Kristen B. Stolka**[3], **Jennifer J. Hemingway-Foday**[3], **Mahamat Abassora**[4], **Paul N'Dom**[4], **Jennifer S. Smith**[5], **Neneh Sallah**[6], **Anne L. Palser**[7], **Paul Kellam**[7,8], **Brandon F. Keele**[2], **Denise Whitby**[1]

[1]Viral Oncology Section, AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

[2]Retroviral Evolution Section, AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

[3]RTI International, Research Triangle Park, North Carolina, USA

[4]SOCHIMIO, Yaoundé, Cameroon

[5]University of North Carolina, Chapel Hill, North Carolina, USA

[6]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK

[7]Kymab Ltd., Babraham Research Campus, Cambridge, UK

[8]Department of Medicine, Division of Infectious Diseases, Imperial College London, London, UK

Correspondence Denise Whitby, AIDS and Cancer Virus Program, Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA. whitbyd@mail.nih.gov.

AUTHOR CONTRIBUTIONS

Study design, performed assays, analyses, drafted figures and co-wrote the article: Vickie A. Marshall; Bioinformatic analyses, drafted figures and significant contributions to the article: Nicholas C. Fisher, Charles A. Goodman, Elena M. Cornejo Castro, Nazzarena Labo and Isabella Liu; Custom library assay design and significant contributions to the article: Paul Kellam, Anne L. Palser; Neneh Sallah; Performed assays: Benjamin M. Holdridge; Sequence alignment bioinformatic pipeline creation and analysis support: Sirish Khanal and Abigail L. Thorpe; Study implementation, biological sample shipments and quality control of clinical data collected: Kristen B. Stolka; Study participant recruitment and management of study implementation and clinical data collection in Cameroon: Paul N'Dom, Mahamat Abassora; Significant contribution to the interpretation and presentation of findings and technical assay support: Brandon F. Keele; Study design and oversight of all study implementations: Jennifer S. Smith; Study design and questionnaire development: Jennifer J. Hemingway-Foday; Study design, co-wrote the article and management of the project: Denise Whitby. The work reported in the article has been performed by the authors, unless clearly specified in the text.

CONFLICT OF INTEREST

Dr. Neneh Salah is currently an employee of GSK and owns company shares. No other authors have COI to declare.

ETHICS STATEMENT

The Cameroon KS case-control study was conducted within the framework of the International Epidemiologic Databases to Evaluate AIDS (IeDEA) Central Africa cohort. The study participants all gave informed consent according to the declaration of Helsinki. The study was approved by the Comité National D'Ethique/Cameroon National Ethics Committee and by the IRBs of the Research Triangle Institute (RTI) International as well as the U.S. National Cancer Institute.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Abstract

In sub-Saharan Africa, Kaposi's sarcoma-associated herpesvirus (KSHV) is endemic, and Kaposi's sarcoma (KS) is a significant public health problem. Until recently, KSHV genotype analysis was performed using variable gene regions, representing a small fraction of the genome, and thus the contribution of sequence variation to viral transmission or pathogenesis are understudied. We performed near full-length KSHV genome sequence analysis on samples from 43 individuals selected from a large Cameroonian KS case-control study. KSHV genomes were obtained from 21 KS patients and 22 control participants. Phylogenetic analysis of the K1 region indicated the majority of sequences were A5 or B1 subtypes and all three K15 alleles were represented. Unique polymorphisms in the KSHV genome were observed including large gene deletions. We found evidence of multiple distinct KSHV genotypes in three individuals. Additionally, our analyses indicate that recombination is prevalent suggesting that multiple KSHV infections may not be uncommon overall. Most importantly, a detailed analysis of KSHV genomes from KS patients and control participants did not find a correlation between viral sequence variations and disease. Our study is the first to systematically compare near full-length KSHV genome sequences between KS cases and controls in the same endemic region to identify possible sequence variations associated with disease risk.

Keywords

Cameroon; human herpesvirus 8; Kaposi's sarcoma; Kaposi's sarcoma-associated herpesvirus; next-generation sequencing; whole genome sequencing

1 | INTRODUCTION

Kaposi's sarcoma-associated herpesvirus (KSHV) is the causative agent of Kaposi's sarcoma (KS), primary effusion lymphoma (PEL), a variant of multicentric Castleman's disease (MCD) and KSHV-inflammatory cytokine syndrome (KICS).[1] KS is a significant public health problem in sub-Saharan Africa (SSA), particularly in the context of generalized HIV/AIDS epidemics.[2,3] Prior to the wide-spread availability of antiretroviral treatments, KS was the most common cancer in men and second most common in women in Uganda, Zimbabwe and Malawi.[4] More recently it was reported that KS incidence has increased and is among the top three most prevalent cancers in both men and women in Johannesburg, South Africa.[5]

KS was first described in 1872 as an indolent disease occurring primarily in men of Mediterranean descent: this clinical-epidemiological variant is now referred to as classical KS. KS has additional recognized variants including African endemic, iatrogenic, AIDS-associated and the recently proposed form, KS in MSM, described in HIV-negative men who have sex with men (MSM).[6] KS is an angioproliferative disease in which spindle cells of endothelial origin are latently infected with KSHV. All types of KS are caused by KSHV infection and are similar pathologically but differ in clinical presentation and course.[7] African endemic KS frequently occurs in patients of younger age, including children and can have a more severe clinical course and poor prognosis as it is often diagnosed in late stages.[8]

KSHV has been classified into subtypes A to F based on the sequence of the variable K1 gene. The distribution of K1 subtypes is associated with race/ethnicity and geography; subtypes A and C are common in Europe, Asia and the United States, while A5 and B are common in SSA. Subtype D is found in indigenous Austronesian populations and subtype E is found in South American Amerindians.[9,10] Subtype F was first described in SSA, particularly in the San people of Botswana and the Gisu people in Uganda but recently has also been reported in MSM in the United States and France.[11–13] Genetic variation in other gene regions has also been reported, notably K15, K12/T0.7, ORF 26, ORF 75 and the KSHV encoded microRNA region, which have been used to study viral recombination.[12,14,15] Within the last few years, advances in next-generation sequencing (NGS) have enabled sequencing of near full-length KSHV genomes. These include 16 KS-derived sequences from AIDS-KS patients in Zambia, 4 non-AIDS KS from the Japanese Miyako Islands, 7 from HIV negative MSM in France and 23 from matched saliva and KS tumors of Ugandan patients.[13,16–18] We have reported near full-length genomes from 46 KSHV infected healthy participants of the Uganda General Population cohort and more recently from nine patients with effusions in the United States.[1,19] To date, however, no studies have compared KSHV genomes from KS cases and controls from the same geographic region, making it difficult to determine if viral genetic variations are associated with disease risk.

We have previously published findings from an AIDS KS case-control study conducted in Cameroon. KSHV seroprevalence was ≥80% in controls; risk factors associated with KS included, minority ethnicity, family history of cancer, HIV infection and factors related to malaria and other parasitic diseases.[20] A subsequent study investigated the relationship between KSHV load in blood and oral fluids, and we[21] recently reported associations between specific HLA alleles and risk for both AIDS and endemic KS in this population.[22]

The current work extends our observations in this well-characterized study to include near full-length KSHV genome sequences circulating in Cameroon using a target enrichment NGS approach.[19] A systematic analysis of viral sequence variation was performed comparing viral genomes from AIDS-KS patients and controls. KSHV genes of noncoding RNAs, including viral microRNAs, were examined in depth as single nucleotide variations (SNVs) have been associated with the risk of KSHV-associated MCD.[23] To our knowledge, this is the first study of near full-length KSHV genomes conducted in a KS endemic region, that compares sequences from KS cases and controls within the same source population.

2 | MATERIALS AND METHODS

2.1 | Study participants

The Cameroon KS case-control study was conducted within the framework of the International Epidemiologic Databases to Evaluate AIDS (IeDEA) Central Africa cohort as reported previously and is described here.[20–22] Briefly, HIV infected and uninfected individuals with KS were recruited through Solidarité Chimiothérapie (SOCHIMIO) and the General Hospital of Yaoundé. KS was confirmed by a senior oncologist after clinical and/or pathological diagnoses. Age and sex matched controls were recruited at the General Hospital or the Military Hospital of Yaoundé, both IeDEA sites.

2.2 | Sample collection and processing

Matched blood and oral fluids samples were collected in PAXgene DNA tubes and in alcohol-based mouthwash respectively, from Cameroonian study participants at enrollment and during follow up visits at 6 and 12 months. All samples were shipped to the Viral Oncology Section (VOS) at the end of each collection period for processing. The processing of samples and qPCR testing has been previously described.[21]

2.3 | Next generation sequencing

Samples were selected based upon KSHV load estimates >1000 copies as measured by qPCR (Table 1). Libraries were prepared using the Agilent SureSelect XT Custom Library Preparation Kit (Agilent Technologies, Santa Clara, California). A custom bait set capturing KSHV and EBV was designed as previously described.[19] Each sample library was labeled with a unique barcode to allow for pooling of samples. Paired-end libraries were sequenced using an Illumina MiSeq Instrument (500 cycle, V2 cartridge: Illumina, Hayward, California).

2.3.1 | Sanger sequencing—Sanger sequencing was used to supplement NGS in some gene regions. Gaps in the K15 N gene region were Sanger sequenced in five samples. Moreover, Sanger sequencing was used to cover small areas of gaps in KSHV coverage as noted in Table S1. The dual infection in sample FNL0985_CM was confirmed with a targeted primer strategy as previously described[1] in five regions: ORF8, vIRF-2, ORF 56, ORF57 and K15, using the original DNA extracted for NGS analysis. In addition, various combinations of primers for the K1 and K15 genes were used to assess the deletions in samples FNL0515_CM and FNL0297_CM, respectively.[12] All primers were designed using Geneious Prime January 1, 2020 software (https://www.geneious.com) and are described in Table S1.

2.4 | KSHV sequence assembly, alignment and VCF generation

2.4.1 | Sequence guided alignment—BBDuk was used to trim raw reads of artifact adapter sequence and bases with PHRED quality score below 32. Reads shorter than 100 bp were removed (BBMap-Bushnell B; https://sourceforge.net/projects/bbmap/). Trimmed reads were initially aligned to GK18 (NC_009333.1) using BWA MEM.[24] Resulting SAM files were processed via Samtools[25,26] and deduplicated with GATK (http://broadinstitute.github.io/picard/). VCF files were generated with Picard to separately identify SNPs and INDELs with respect to the GK18 reference sequence via GATK. SNPs failing to pass Variant Filtration expression “QD <2.0 ||FS >60.0 ||MQ <40.0 ||MQRankSum < −12.5||ReadPosRankSum < −8.0” and INDELs failing to pass “QD <2.0 ||FS > 200.0 ||ReadPosRankSum < −20.0” were removed. SNP and INDEL VCFs were combined and validated for the final output.[27] BC-1 (U75698.1) and ZM128 (KT271467) sequences were subsequently used to correct alignments for K15 M and N subtype gene regions respectively. Sequences were manually inspected for misalignments in Geneious and five areas of repetitive sequence were masked including positions 24 230 to 25 045, 29 927 to 30 055,118 229 to 113 914, 124 784 to 126 456 and 137 169 to 137 969, referencing the GK18 reference genome (NC_009333.1).[26,28] Study sequences with poor read depth, insufficient

coverage and/or high overall percentages of non-calls were removed from the analysis. We used a combination of median coverage of ×30 and reference bases coverage as >98% to determine which samples passed QC (Table S2). Consensus FASTA files were exported from Geneious.

2.5 | Phylogenomic analysis

Consensus KSHV sequences with complete coverage outside of repeat regions were aligned with 16 available full-length genomes published in GenBank as of April 2019 in Geneious, using the MAFFT algorithm with default settings.[26]

The resulting alignment was used to infer phylogenomic trees via the neighbor-joining method visualized using Mega 6 software (Figure 1A,B).[29] Maximum likelihood phylogenetic trees were constructed for both the complete nucleotide alignment and a subalignment lacking K1 and K15 regions using IQtree version 1.6.12 with 1000 bootstrap replicates and evolution models determined empirically via the MFP setting[30] (Figure S1). The KSHV K1 and K15 gene sequences were aligned with representative published sequences to determine the subtype (Figure 1A,B).

A SplitsTree of the study sequences was created using the default settings (Figure 1C). Recombination was assessed using the Pairwise Homoplasy Index (PHI) text implemented in SplitsTree v4.15.1.[31] Recombination was further assessed using Simplot software, with results visualized using Bootscan.[32,33] The analysis included the reference sequences GK18 (NC_009333.1), BC-1 (U75698.1), FNL014_CM (MN419227.1), ZM095 (KT271456.1), Japan 1 (LC200589.1), VG-1 (unpublished Whitby laboratory), BCBL-1 (HQ404500.1) and JSC-1 (GQ994935.1) (Figure S2).

2.6 | KSHV percent similarity by gene region

Subalignments of all coding regions (CDS) and functionally described noncoding regions were obtained, and pairwise distance matrices were calculated and exported from Geneious. Boxplots reflecting the distribution of pairwise distances for select regions across the KSHV genome were produced in R as shown in Figure 2A (R Core Team [2020]. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/). Clustered heatmaps were constructed using pheatmap version 1.0.12 in R to visualize similarities between individual sample pairings, and to identify groupings of samples based on sequence similarity while searching for evidence of K1-and/or K15-derived designations therein (Figure S5).

2.7 | Circos analysis

An annotated KSHV genome map was constructed with the GK18 reference genome (NC_009333) highlighting regions not covered by our custom bait set (Figure 3A). The positions of all identified SNVs and indels detected in the study genomes as well as the linkage disequilibrium (LD), quantified as the correlation coefficient R^2 in PLINK v v1.9[34] (https://zzz.bwh.harvard.edu/), are visually represented in Figure 3B.[35]

3 | ASSOCIATION BETWEEN SNVS AND CASE/CONTROL STATUS

3.1 | Principal component analysis

Principal component analysis (PCA) of sequences from HIV+ cases and controls was performed by generating eigenvalues from a VCF file using PLINK v1.9 to identify potential sample clustering.[36] After the identification of variants with a minor allele frequency >0.05, 1217 variants and 41 individuals (20 cases/21 controls) were included in the analysis. The average identity was computed in PLINK, and one sample from each closely related pair (PI_HAT >0.2) was further excluded. To ensure that the principal components (PCs) represent the genome-wide structure, SNVs were further pruned by LD (>0.1). The top five pairs of PCs across 35 samples (19 cases/16 controls) and 147 SNVs were plotted in R with each sample colored by case or control status.[36]

3.2 | Test for proportion

The proportion of sequences from cases and controls with SNVs at any given position is represented in Figure 4. To infer whether SNVs were more common in KS vs control sequences, we performed a parametric test for equality of proportion for each position without correcting for multiple comparison; the relevant probability is also represented visually in Figure 4.

3.3 | Analysis of noncoding RNA regions

Polymorphisms within the microRNA coding regions were mapped and KS case/control status as well as KSHV K1 subtype for each were visualized using Lolliplot software in conjunction with TrackViewer (https://github.com/pbnjay/lollipops).[37] In addition, a parametric test for equality was performed, as described above, to determine if any polymorphisms were associated with KS (Figure S3).

4 | RESULTS

4.1 | KSHV phylogenetic and genome analysis

Near full-length KSHV sequences were successfully obtained from 43 of 52 participants using a target enrichment NGS approach from whole blood or oral fluids with detectable viral load (VL) (Table 1). Nine samples, which tended to have lower KSHV VL, had insufficient read depth coverage and were not included in downstream analyses as noted in Table S2. Although samples with KSHV loads higher than 10 000 copies were more likely to result in successful near full-length KSHV genome coverage, it was noted that in two instances oral fluid DNA for FNL0432 and FNL068, did not amplify as well, despite having VL >10 000 copies. Reasons for failure to produce near full-length genomes are unknown, but generally oral fluid DNA is more fragmented than DNA extracted from blood which can affect performance. In addition, the EBV load for sample FNL0681 was exceedingly high, which could adversely affect KSHV read depth coverage as our custom bait set simultaneously captures both viruses. The range of median coverage for samples that passed QC was 50 to 4252 X. Of the 43 participants included, 21 were KS cases (20 AIDS-KS and 1 endemic KS) and 22 controls (21 HIV+, 1 HIV—). In all but three KS cases, KSHV genome sequences were obtained from whole blood while for all but one control, oral fluids

were sequenced. The genome sequences for three samples could not be resolved due to mixed KSHV infections, resulting in a final 40 genome sequences available for analysis.

4.1.1 | K1 gene—Two KSHV K1 gene subtypes were prevalent in our study population, A5 (24/40, 60%) and B1 (14/40, 35%). One sample from a control participant, FNL0173_CM, had a B2 subtype (Figure 1A). A complete deletion of the K1 gene which extended into the ORF4 CDS was identified in FNL0515_CM, a KS case. The deletion was confirmed by NGS in both a blood and an oral fluid sample sequenced 4 months apart. Attempts to Sanger sequence across this region using custom primers, as shown in Table S1, produced a product that did not extend into the gap in coverage. The Sanger product failed to confirm 24 nucleotides of KSHV sequence just downstream of the gap. To rule out reference guided alignment bias, this sample was also assembled de novo producing a genome that also had a complete K1/partial ORF4 gap sharing 96.3% overall identity with the reference guided aligned sequence (data not shown). Amid the 39 remaining sequences, similar numbers of KS cases and controls were determined to have A5 (11/18, 61% vs 13/20, 65%) or B1 (7/18, 39% vs 7/20, 35%) subtypes (B2 subtype control excluded).

4.1.2 | K15 gene—All three K15 genotypes are represented, with the predominant P subtype most abundant (25/40, 62.5%), followed by the minor M (9/40, 22.5%) and the N (5/40, 12.5%) as shown in Figure 1B. The sequence FNL0297_CM, isolated from a control participant, carried an apparent partial deletion of the K15 CDS. The gap was initially thought to be a low coverage sequencing artifact. Primers were designed spanning the missing region referencing the NC_009333 genome and FNL0297 K15 specific forward primers to Sanger sequence into the gap. Sanger sequence ultimately proved inconclusive, and this sequence was also assembled de novo to rule out reference guided alignment bias. A de novo alignment confirmed the K15 deletion, displaying 98.05% identity to the reference guided consensus sequence (data not shown). In addition, the sequence of control sample FNL0173_CM, a K15 M subtype, has polymorphisms within the ORF75/K15 gene overlap that are clearly the K15N subtype as shown in Figure S6. This suggests a recombination event occurred near the end of the overlap of the ORF75 and K15 genes between nucleotides 134 569 and 134 595 (referencing NC_009333.1). Among KS cases 73.7% (14/19) were of the P subtype, 10.5% (2/19) of the M and 15.8% (3/19) of the N subtype. Controls were also predominantly P subtype (55%, 11/20), followed by M (35%, 7/20) and N (10%, (2/20).

4.1.3 | microRNA coding region—The KSHV microRNA encoding region includes the K12 (kaposin) gene and the intragenic region between the K12 and ORF71 genes. Many SNVs previously identified in the microRNA region, common in KSHV sequences from Africa, were observed in the current study.[38] However, novel sequence variants were noted within the K12 gene including a 11 bp deletion, located within the T0.7 product of K12, observed in sequences from three control and two KS cases. While polymorphisms were frequent, none were statistically associated with KS although some were more frequently observed in genomes of either K1 subtype A5 or B (Figure S3).

4.1.4 | KSHV genome analysis and gene diversity—A SplitsTree of 40 Cameroon samples, excluding the dual infections, and 16 published genomes shows a high degree of

conflicting phylogenetic signals (Figure 1C). Five regions of repeats, representing <6% of the viral genomes as described in the sequence alignment section of Materials and Methods, were not covered by our custom baits, and were masked during analysis (Figure 3A).

The genetic diversity of the CDS was analyzed by comparing the Cameroon KSHV A5 and B subtype sequences, to the canonical C subtype GK18 reference. The lowest nucleotide identity across all sequences was observed in known variable genes K1 and K15, ≥20% and 50%, respectively. Outside of these regions, overall variation was low, with 99.5% to 98.5% identity excluding samples FNL0515_CM and FNL0297_CM that have significant deletions in the K1 and K15 gene regions. Many genes within the central portion of the KSHV genome were conserved between study sequences in comparison to each other and the reference genome. Conserved genes include ORF36 (99.8%), K5 (99.75%), ORF35, ORF40 and ORF43 (all 99.6%) as noted in Figure 2. KSHV genes with notably higher variation and unique polymorphisms include the ORF47 (90%), K8 (95%) and K4.2 (98.6%) gene regions (Figure 2B).

4.2 | Single nucleotide variants in KS cases and controls

When compared to GK18, we identified 1184 SNVs arising in at least 25% of sequenced samples. Of these, 376 SNVs were non-synonymous mutations. SNVs were distributed across the whole KSHV genome, with specific patterns of LD noted in Figure 3B. PCA did not support the segregation of cases and controls based upon the unique variants observed in these samples (data not shown). In addition, parametric tests for the equality of proportion determined that no polymorphisms distinguished KS or control sequences as the SNVs identified were observed with comparable proportion in both cases and controls (Figure S4B). Additional multivariate analysis was performed examining both KSHV K1 and K15 subtype and determined that neither were associated with KS risk (data not shown).

4.3 | Evidence of multiple infections

Samples of three individuals, FNL0454_CM, FNL0578_CM and FNL0985_CM, showed evidence for the presence of at least two divergent KSHV genomes. FNL0454_CM and FNL0578_CM contained reads specific to both the A5 and B K1 subtypes, as illustrated in Figure 5A,B. The dual infection pattern of FNL0985_CM is only apparent outside of the K1 gene region, as indicated by two different K15 subtypes, P and N, and distinct patterns of reads supporting different subtypes in genes within the central portion of the genome. The dual infection in FNL0985_CM was confirmed by Sanger sequencing using strain specific primers in the ORF8, ORF57 and K15 gene regions (Figure 5C; Table S1). Unfortunately, insufficient material remained for samples FNL0578_CM or FNL0454_CM to confirm these observations by Sanger sequencing.

4.4 | Recombination analysis

Simplot software analysis indicated that the study sequences were highly recombinant. Cameroonian genomes were compared to eight KSHV published genomes of all available KSHV K1 and K15 subtypes, excluding the K1 F subtype which was published after our analysis. Unique fragmentation profiles were generated illustrating potential recombination sites visualized using Bootscan software. The Cameroonian sequences most closely

resembled reference sequences of A5/B KSHV K1 subtypes, as expected (Figure S2). A SplitsTree was created using only the 40 Cameroon genomes with single KSHV infections. This resulted in a network of three splits driven by the highly variable K15 gene. FNL0297, which has a substantial K15 deletion, is positioned in the central node of the SplitsTree despite having an otherwise intact KSHV genome thus emphasizing the strong influence of K15. A Phi test of recombination determined using the phylogenetic network created in SplitsTree, showed strong evidence of recombination between circulating KSHV strains in Cameroon.

5 | DISCUSSION

A total of 40 near full-length KSHV genomes were successfully obtained from a Cameroon KS case control study using a target enrichment bait design previously reported.[19,39] Consistent with previous reports from Cameroon, the KSHV sequences detected in this population were A5 or B K1 subtypes.[40] One sample, FNL0515_CM, could not be subtyped due to a deletion of the K1 gene region that extended into the downstream ORF4 CDS. We observed that the distinctive pattern of polymorphisms, used to distinguish KSHV K1 subtypes, were not limited to K1 but extended into the ORF4 gene. This distinctive region of variability encompasses the region missing in FNL0515_CM, a KS case. The apparent K1 gene deletion in this KS case sequence implies that the K1 gene, a recognized KSHV oncoprotein, may not be strictly required for KS development[41] (Figure S4A). In addition, the substantial partial deletion in ORF4 includes the start codon suggesting that the encoded protein is unlikely to be produced.

All three K15 subtypes, P, M and N were identified in the study (Figure 1B). One genome, FNL0173_CM, has a recombination event, at the beginning of the overlap between the ORF75 and K15 genes (Figure S6). Similar recombination events of ORF75/K15 have been previously reported.[1,14] As the breakpoint occurred after the end of the CDS, it is not likely to affect processing of the either protein. The K15 gene was the strongest driver of KSHV genome topology as shown in a SplitTree of 40 samples, excluding the three dual infections (Figure S2B) and also in IQTrees that visualizes the contribution of the K1 and K15 regions in topologies in which K1, K15 or both K1 and K15 have been omitted from the near full-length sequences (Figure S1A-D).

The phylogenomic analysis of the near full-length KSHV sequences did not precisely recapitulate the distinct splits pattern observed by Sallah et al used for the reclassification of KSHV K1 subtypes[19] (Figures 1C and S2D) although major branching patterns remained consistent. Our analysis did confirm previous observations[1,16] in that, although K1, and particularly K15, were strong drivers of the branching patterns observed in the SplitsTree, variation within the central portion of the KSHV genome also contributes to whole genome phylogeny (Figure 1C). This supports the merit of considering full-length genomes, when available, in determining subtype designations of KSHV strains. By using whole genomes sequences and not just highly variable genes, the distinctive patterns of variation observed across the viral genomes, as shown in Figure S4, could then be considered. A recent study of whole EBV genomes noted that geography, as well as recombination, were essential considerations in subtype designation.[42] In order to assess recombination patterns in KSHV,

the distribution and linkage among SNVs across all genomes was visualized in Figure 3B. The high level of LD across the genome suggests that isolated variations may have wider implications for future association analyses and the identification of recombinant loci.

A high conservation of 94.6% to 99.6% was observed within genes in the central portion of the KSHV genome. However, a high frequency of nonsynonymous mutations relative to GK18 was noted in the K4.2, ORF47 and K8 genes which could alter function or viral phenotype (Figure 2B). Mutations in K4.2, encoding an pERP-1 binding protein which regulates immunoglobulin secretion and calcium homeostasis, have been previously reported[16]; our data included 13 samples with a frame shift introducing seven amino acid changes, followed by a premature stop codon, potentially resulting in a truncated protein.[43] In addition, two samples (FNL0531_CM and FNL0346_CM) had frame shifts due to nucleotide insertions resulting in significant amino acid variations near the 3′ end of the 4.2 coding region. In KS samples FNL0515_CM and FNL0461_CM, we identified a 15 base pair deletion and a high degree of nonsynonymous mutations within the ORF47 CDS, which encodes the virus glycoprotein L. A search of available KSHV genomes in GenBank found these variations in sequences from Japan, Uganda and France.[13,17,18,44] The mutations do not appear to alter the start and stop codon of the protein but could potentially have functional implications. In K8, encoding for a bZIP protein analogous to the EBV Zta protein, 34 of 40 sequences had a single nucleotide deletion resulting in a series of nonsynonymous amino acid changes followed by a premature stop codon. The phenotypic consequences of the sequence variants observed in these genes are beyond the scope of our study and merit further investigation.

Previous studies have suggested relationships between risk, severity or outcomes of KS and specific KSHV variants; however, many studies have primarily focused on viral sequences from KS patients and, with the exception of a recent MSM study in France, often focused on exclusively highly variable genes like K1.[13,45–47] In this article, we have the opportunity to examine KSHV sequence variations, including coding and noncoding regions from both AIDS-KS cases and controls, from the same source population. After extensive analyses, we were unable to observe specific genetic variations in the KSHV genome that were associated with disease risk. PCA analysis could not distinguish KS case and control samples or K1 subtypes (data not shown). A parametric test of equality further corroborated this finding, which failed to show sequence variations distinguishing KS cases and controls (Figure 4). In a previous study, we have shown unique KSHV microRNA SNVs associated with mature microRNA expression differences and increased risk of MCD.[48] However, as all study samples share similar sequence variations across this region, we did not find an association between microRNA sequence variation to KS risk. This finding contrasted with our study of a European AIDS-KS study, comprised of samples from five countries of varying KSHV K1 subtypes, in which we noted higher frequency of SNVs within the microRNA region in cases compared to controls.[49] In a similar manner, we also examined the noncoding RNA PAN region in which SNVs also clustered equally among cases and controls (data not shown).

There are obvious study limitations that may impact our findings. As the study design was essentially cross-sectional, some control participants could have later developed KS;

however, given the cumulative incidence of KS in HIV-infected Cameroonian and in the general population, it is unlikely that substantial misclassification occurred in our study. The study was not powered to investigate the effect of HIV on KSHV sequence variation and KS risk as few samples from HIV negative individuals qualified for NGS. Additional studies examining KSHV genomes from HIV negative individuals would be informative. The custom baits used were designed to simultaneously capture both KSHV and EBV. As many of the study samples had extremely high EBV copies, particularly in oral fluids (data not shown), this impacted the read depth coverage obtained from nine samples, as shown in Table S2. Only samples with KSHV loads greater than 10 000 copies reliably achieved sufficient coverage of the KSHV genome, limiting the number of specimens in our study, and potentially introducing selection bias into our analysis.[50] Based on the technical limitations of short read sequencing, the repetitive areas in close proximity to K4, K7, the microRNA cluster and K15 as well as the internal repeats within ORF73 and the long terminal repeat between K1 and K15 were not covered by our custom bait design. Thus, the variation in these regions and their possible contribution to pathogenicity could not be assessed. Although the sequences characterized by the study reflect the viral variants circulating in this population, they were obtained from whole blood or oral fluids and not from KS biopsies. This may explain why large structural mutations, recently described by Santiago et al, were not observed within our sample set.[18] While the study was balanced between cases and controls, 40 genomes may not be sufficient to detect low frequency KS risk associations. Expanded studies, utilizing long read sequencing to capture repetitive regions of the viral genome, are needed to investigate the contribution of sequence variations more adequately to KS risk.

We observed within-sample sequence heterogeneity in three samples consistent with the presence of two KSHV genomes, as identified by distinct K1 or K15 subtypes (Figure 5). The sequence heterogeneity was apparent across the whole genome and not only observed in highly variable KSHV gene regions. Whether a haplotype analysis from short read data would resolve the viral diversity observed in the dually infected samples across the genome fragments is being investigated.[50] It is possible that many multiple infections are undetected due to inherent detection limitations in NGS protocols including the dependence upon high VL of each genome variant within in the sample and the read depth coverage available for analysis. These observations support our recently reported dual KSHV infection in an PEL patient, also from Cameroon, which we identified using a shot gun sequencing approach.[1]

In addition to unique sequence variations and observation of multiple infections, we examined the genomes for evidence of recombination. The results suggest that the Cameroon samples are highly fragmented and most closely related to A5 and B references as expected.[1,16] These observations were further supported by a PHI test that provided strong statistical evidence of recombination.

Our study included direct comparisons of near full-length KSHV genomes from both KS case and control participants from the same endemic KS region, finding no evidence of associations between individual SNVs and disease. The joint or cumulative effect of numerous polymorphisms is beyond the scope of our study but highlight the need to further characterize the contributions of viral sequence variations to KSHV pathogenesis.

Moreover, our current observations of highly recombinant KSHV sequences and mixed KSHV infections, may have implications for future KSHV vaccine strategies, particularly in KSHV endemic regions. Further, our observations highlight the phylogenetic analysis of geographically distinct viral genomes as an important epidemiological tool. The authors of a recent article on EBV addressing phylogenetic challenges emphasize the need to increase the number of viral genomes available for analysis with a focus on characterizing world-wide sequence diversity.[42] To that end, our study provides an additional 40 near full-length KSHV sequences to the research community, from a KSHV endemic region, to further promote our understanding of KSHV epidemiology and evolution.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank Carolyn Reid, Leslie Lipkey, Laura Newman and Agatha Macairan for technical assistance as well as Joseph Meyer for graphical design expertise. In addition, we are especially grateful to the study participants.

FUNDING INFORMATION

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. 75N91019D00024 (FFRDC contract from 8/31/19 to present) and HHSN261200800001E (OTS contract from 9/26/08 to 8/30/19).

DATA AVAILABILITY STATEMENT

Forty near full-length KSHV sequences are publicly available in GenBank with accession numbers OL829860 to OL829899. Further details and other data that support the findings of our study are available from the corresponding author upon request.

Abbreviations:

CDS	coding region
HHV8	human herpesvirus 8
IeDEA	International Epidemiologic Database to Evaluate AIDS
KICS	KSHV-inflammatory cytokine syndrome
KS	Kaposi’s sarcoma
KSHV	Kaposi’s sarcoma-associated herpesvirus
LD	linkage disequilibrium
MCD	multicentric Castleman’s disease
MSM	men who have sex with men
NGS	next-generation sequencing

PC	principal components
PCA	principal component analysis
PEL	primary effusion lymphoma
PHI	pairwise homoplasy index
SNVs	single nucleotide variations
SOCHIMIO	Solidarité Chimiotherapie
SSA	sub-Saharan Africa
VCF	variant call format
VOS	Viral Oncology Section

REFERENCES

1. Cornejo Castro EM, Marshall V, Lack J, et al. Dual infection and recombination of Kaposi sarcoma herpesvirus revealed by whole-genome sequence analysis of effusion samples. Virus Evol. 2020;6: veaa047.
2. Parkin DM, Hammerl L, Ferlay J, Kantelhardt EJ. Cancer in Africa 2018: the role of infections. Int J Cancer. 2020;146:2089–2103. [PubMed: 31254479]
3. Sankaranarayanan R, Swaminathan R, World Health Organization, International Agency for Research on Cancer. Cancer Survival in Africa, Asia, the Caribbean and Central America. Lyon, France: International Agency for Research on Cancer, World Health Organization; 2011.
4. Cesarman E, Damania B, Krown SE, Martin J, Bower M, Whitby D. Kaposi sarcoma. Nat Rev Dis Primers. 2019;5:9. [PubMed: 30705286]
5. Chen WC, Singh E, Muchengeti M, et al. Johannesburg cancer study (JCS): contribution to knowledge and opportunities arising from 20 years of data collection in an African setting. Cancer Epidemiol. 2020; 65:101701.
6. Denis D, Seta V, Regnier-Rosencher E, et al. A fifth subtype of Kaposi's sarcoma, classic Kaposi's sarcoma in men who have sex with men: a cohort study in Paris. J Eur Acad Dermatol Venereol. 2018;32: 1377–1384. [PubMed: 29377280]
7. Dupin N Update on oncogenesis and therapy for Kaposi sarcoma. Curr Opin Oncol. 2020;32:122–128. [PubMed: 31815777]
8. Orem J Cancer prevention and control: Kaposi's sarcoma. Ecancermedicalscience. 2019;13:951. [PubMed: 31552124]
9. Hayward GS. KSHV strains: the origins and global spread of the virus. Semin Cancer Biol. 1999;9:187–199. [PubMed: 10343070]
10. Whitby D, Marshall VA, Bagni RK, et al. Genotypic characterization of Kaposi's sarcoma-associated herpesvirus in asymptomatic infected subjects from isolated populations. J Gen Virol. 2004;85:155–163. [PubMed: 14718630]
11. Kajumbula H, Wallace RG, Zong JC, et al. Ugandan Kaposi's sarcoma-associated herpesvirus phylogeny: evidence for cross-ethnic transmission of viral subtypes. Intervirology. 2006;49:133–143. [PubMed: 16428889]
12. Marshall V, Parks T, Bagni R, et al. Conservation of virally encoded microRNAs in Kaposi sarcoma: associated herpesvirus in primary effusion lymphoma cell lines and in patients with Kaposi sarcoma or multicentric Castleman disease. J Infect Dis. 2007;195:645–659. [PubMed: 17262705]
13. Jary A, Leducq V, Desire N, et al. New Kaposi's sarcoma-associated herpesvirus variant in men who have sex with men associated with severe pathologies. J Infect Dis. 2020;222:1320–1328. [PubMed: 32282911]

14. Poole LJ, Zong JC, Ciufo DM, 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;73:6646–6660. [PubMed: 10400762]
15. Zong J, Ciufo DM, Viscidi R, 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: 119–148. [PubMed: 11595592]
16. Olp LN, Jeanniard A, Marimo C, West JT, Wood C. Whole-genome sequencing of Kaposi's sarcoma-associated herpesvirus from Zambian Kaposi's sarcoma biopsy specimens reveals unique viral diversity. J Virol. 2015;89:12299–12308. [PubMed: 26423952]
17. Awazawa R, Utsumi D, Katano H, 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;216:850–858. [PubMed: 28968717]
18. Santiago JC, Goldman JD, Zhao H, et al. Intra-host changes in Kaposi sarcoma-associated herpesvirus genomes in Ugandan adults with Kaposi sarcoma. PLoS Pathog. 2021;17:e1008594.
19. Sallah N, Palser AL, Watson SJ, et al. Genome-wide sequence analysis of Kaposi sarcoma-associated herpesvirus shows diversification driven by recombination. J Infect Dis. 2018;218:1700–1710. [PubMed: 30010810]
20. Stolka K, Ndom P, Hemingway-Foday J, et al. Risk factors for Kaposi's sarcoma among HIV-positive individuals in a case control study in Cameroon. Cancer Epidemiol. 2014;38:137–143. [PubMed: 24631417]
21. Labo N, Marshall V, Miley W, et al. Mutual detection of Kaposi's sarcoma-associated herpesvirus and Epstein-Barr virus in blood and saliva of Cameroonians with and without Kaposi's sarcoma. Int J Cancer. 2019;145:2468–2477. [PubMed: 31265124]
22. Cornejo Castro EM, Morrison BJ, Marshall VA, et al. Relationship between human leukocyte antigen alleles and risk of Kaposi's sarcoma in Cameroon. Genes Immun. 2019;20:684–689. [PubMed: 31105266]
23. Ray A, Marshall V, Uldrick T, et al. Sequence analysis of Kaposi sarcoma-associated herpesvirus (KSHV) microRNAs in patients with multicentric Castleman disease and KSHV-associated inflammatory cytokine syndrome. J Infect Dis. 2012;205:1665–1676. [PubMed: 22448005]
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–1760. [PubMed: 19451168]
25. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–2079. [PubMed: 19505943]
26. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–780. [PubMed: 23329690]
27. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491. [PubMed: 21478889]
28. Kearse M, Moir R, Wilson A, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–1649. [PubMed: 22543367]
29. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30: 2725–2729. [PubMed: 24132122]
30. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–274. [PubMed: 25371430]
31. Huson DH. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics. 1998;14:68–73. [PubMed: 9520503]
32. Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol. 1999;73:152–160. [PubMed: 9847317]
33. Strimmer K, Forslund K, Holland B, Moulton V. A novel exploratory method for visual recombination detection. Genome Biol. 2003; 4:R33. [PubMed: 12734013]

34. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7. [PubMed: 25722852]
35. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–1645. [PubMed: 19541911]
36. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–575. [PubMed: 17701901]
37. Ou J, Zhu LJ. trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data. Nat Methods. 2019; 16:453–454. [PubMed: 31133757]
38. Han SJ, Marshall V, Barsov E, et al. Kaposi's sarcoma-associated herpesvirus microRNA single-nucleotide polymorphisms identified in clinical samples can affect microRNA processing, level of expression, and silencing activity. J Virol. 2013;87:12237–12248. [PubMed: 24006441]
39. Palser AL, Grayson NE, White RE, et al. Genome diversity of Epstein-Barr virus from multiple tumor types and normal infection. J Virol. 2015;89:5222–5237. [PubMed: 25787276]
40. Betsem E, Cassar O, Afonso PV, Fontanet A, Froment A, Gessain A. Epidemiology and genetic variability of HHV-8/KSHV in Pygmy and Bantu populations in Cameroon. PLoS Negl Trop Dis. 2014;8:e2851. [PubMed: 24831295]
41. Wen KW, Wang L, Menke JR, Damania B. Cancers associated with human gammaherpesviruses. FEBS J. 2021. doi:10.1111/febs.16206
42. Zanella L, Riquelme I, Buchegger K, Abanto M, Ili C, Brebi P. A reliable Epstein-Barr virus classification based on phylogenomic and population analyses. Sci Rep. 2019;9:9829. [PubMed: 31285478]
43. Wong LY, Brulois K, Toth Z, et al. The product of Kaposi's sarcoma-associated herpesvirus immediate early gene K4.2 regulates immunoglobulin secretion and calcium homeostasis by interacting with and inhibiting pERP1. J Virol. 2013;87:12069–12079. [PubMed: 23986581]
44. Osawa M, Mine S, Ota S, et al. Establishing and characterizing a new primary effusion lymphoma cell line harboring Kaposi's sarcoma-associated herpesvirus. Infect Agent Cancer. 2016;11:37. [PubMed: 27536332]
45. Mancuso R, Biffi R, Valli M, et al. HHV8 a subtype is associated with rapidly evolving classic Kaposi's sarcoma. J Med Virol. 2008;80:2153–2160. [PubMed: 19040293]
46. Isaacs T, Abera AB, Muloiwa R, Katz AA, Todd G. Genetic diversity of HHV8 subtypes in South Africa: A5 subtype is associated with extensive disease in AIDS-KS. J Med Virol. 2016;88:292–303. [PubMed: 26174882]
47. Bellocchi MC, Svicher V, Ceccherini-Silberstein F. HHV-8 genetic diversification and its impact on severe clinical presentation of associated diseases. J Infect Dis. 2020;222:1250–1253. [PubMed: 32282916]
48. Marshall VA, Labo N, Sztuba-Solinska J, et al. Polymorphisms in KSHV-encoded microRNA sequences affect levels of mature viral microRNA in Kaposi sarcoma lesions. Oncotarget. 2018;9:35856–35869.
49. Marshall V, Martro E, Labo N, et al. Kaposi sarcoma (KS)-associated herpesvirus microRNA sequence analysis and KS risk in a European AIDS-KS case control study. J Infect Dis. 2010;202:1126–1135. [PubMed: 20715927]
50. Cudini J, Roy S, Houldcroft CJ, et al. Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. Proc Natl Acad Sci U S A. 2019;116:5693–5698. [PubMed: 30819890]

What's new?

Kaposi sarcoma-associated herpesvirus (KSHV), the cause of Kaposi sarcoma (KS), is classified into different subtypes based on genetic variation. Little is known, however, about whether variations in KSHV genomes differ among KS patients in the same geographic region. Here, no link was found between viral sequence variations and KS risk, based on comparison of KSHV genomes from KS patients in the same population. Several participants exhibited evidence of at least two divergent KSHV genomes, indicating dual infection with distinct KSHV subtypes. The observation of highly recombinant KSHV sequences and KSHV coinfection contributes to improved understanding of KSHV epidemiology.

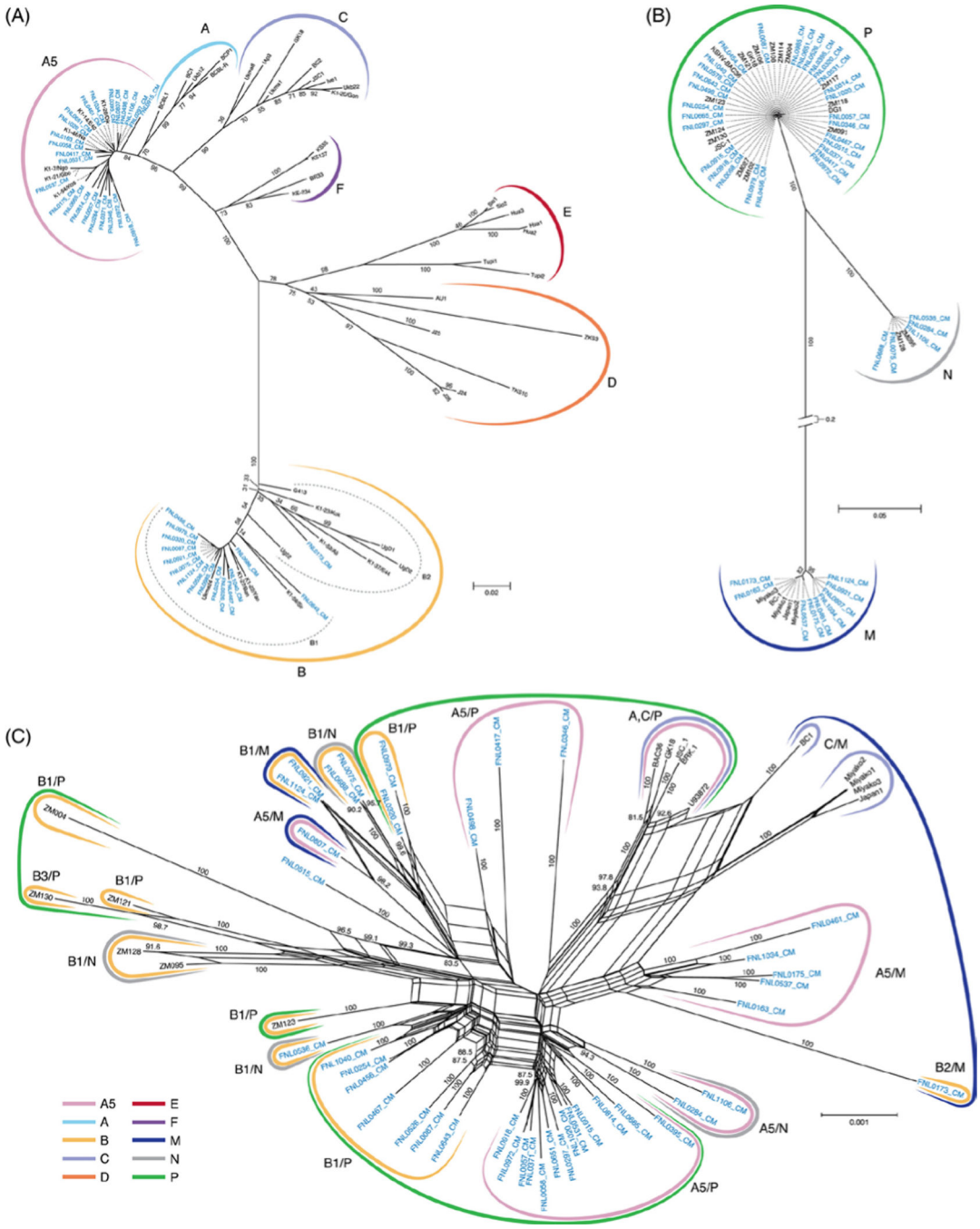


FIGURE 1.

KSHV subtyping (current study participants in blue) with K1 and K15 subtypes indicated by color. (A) K1 neighbor-joining bootstrapped phylogenetic tree created in Mega 6. The analysis consisted of 40 Cameroon and 38 published amino acid sequences. All Cameroon samples were KSHV K1 subtypes A5 or B as indicated (B) K15 neighbor-joining bootstrapped phylogenetic tree constructed in Mega 6. The analysis included 65 nucleotide samples with 777 positions making up the final dataset. Most Cameroon samples were determined to be the predominant K15 P subtype, although less common M and N subtypes

were also observed. (C) Whole-genome splits network of 40 study and 16 published sequences created using SplitsTree 4. A total of approximately 130 000 positions were included in the final data set with bootstrap values over 80 reported. Repeat regions were removed from the analysis. The KSHV full and partial genomes used for phylogenetic analyses included: GK18 (NC_009333), BC-1 (U75698.1), Japan/Miyako sequences (LC200587.1-LC200589.1), BCBL-1 (HQ404500.1), JSC-1 (MK143395.1), KSHV-BAC36 (HQ404500.1), BrK.219 (KF588566.1), BCBL-1 (MT936340.1), Zambian sequences ZM007-ZM130 (KT271453-KT271468), K1–7/Ngo (AF178779), K1–20/Gon (AF178789), K1–21/Gbo (AF178790), K1–22/Yan (AF178791), K1–23/Kok (AF178792), K1–27/Ban (AF178796), K1–31/Ren (AF178798), K1–54/Kos (AF171059), K1–59/Sir (AF178824), K1–34/E40 (AF178801), K1–28/Djk (AF178797), K1–46/Na (AF171058), K1–37/E44 (AF178804), K1–52/Ali (AF178818), Tupi 1 (AF220292.1), Tupi 2 (AF220293.1), Hua1 (AY329026.1), Hua2 (AY329027.1), Hua3 (AY329026.1), Sio1 (AY329025.1), Sio2 (AY329024.1), AU1 (AF151687.1), ZKS3 (AF133044.1), TKS10 (MK176598.1), J24 (AF278844.1), J25 (AF278845.1), J26 (AF278846.1), G413 (AF130262.1), KE-234 (FJ884616.1), BR33 (KT215106.1), KS137 (KP997134.1), KS55 (KP997076.1), UKma24 (AF130301.1), Iap3 (AF130271.1), BCBL-R (AF133038.1), BCP-1 (AY787132.1), Ukma1 (AF130300.1), Ukb22 (AF130298.1), Ive1 (AF130286.1), UgD1 (AF130292.1), UgD2 (AF130293.1), Ug52 (AF130290.1) and BC-2 (AF133042.1)

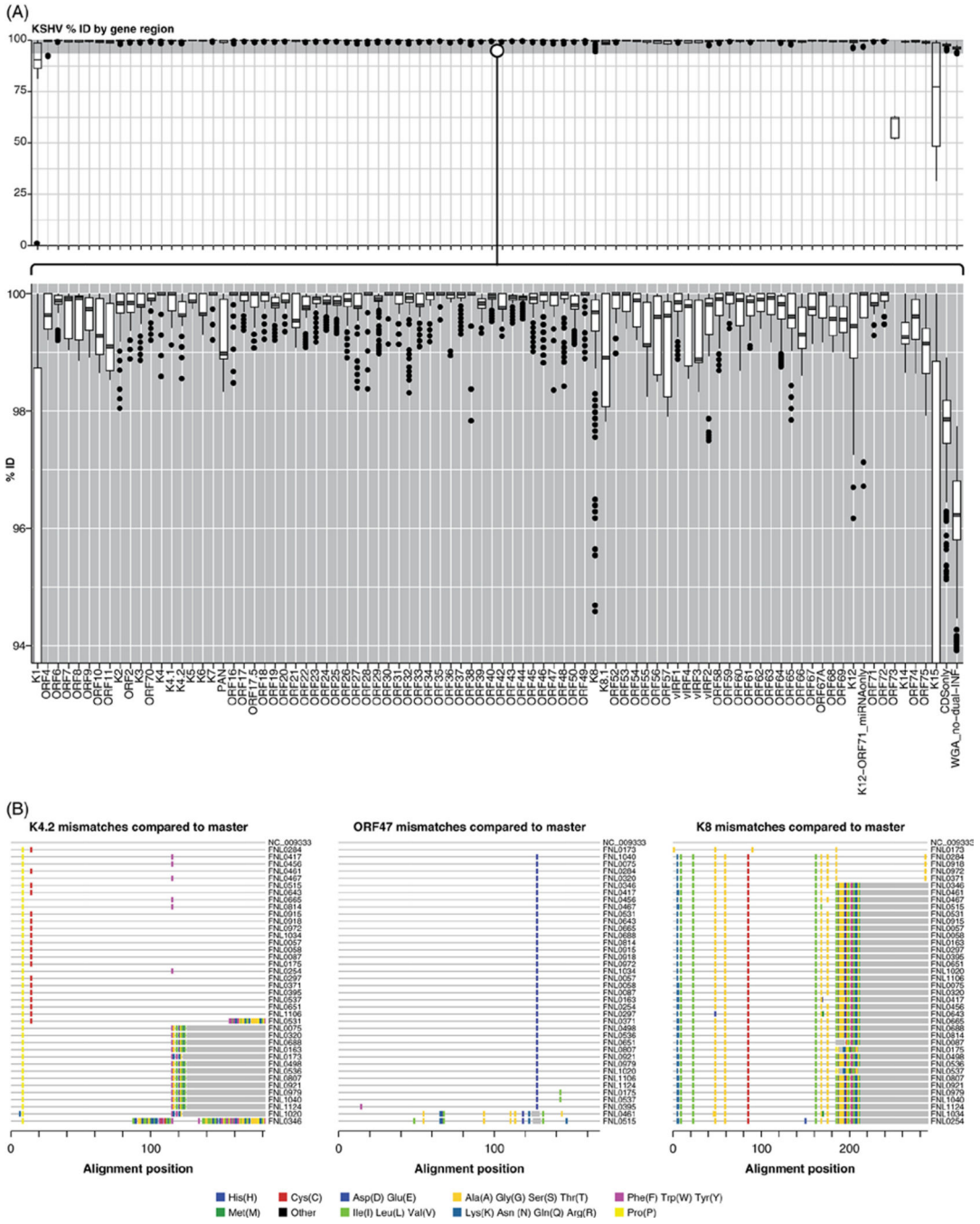


FIGURE 2.

KSHV %ID by gene region: (A) Boxplots reflect the distribution of pairwise distances calculated for sample pairs in selected regions across the KSHV genome. Distances are based on a consensus alignment and includes current study sequences and the KSHV GK18 reference NC_009333.1. As expected, K1, ORF73 (repeat regions masked) and K15 are highly divergent regions. Variation outside these gene regions, in the central portion of the viral genome, ranged from 0.2% to 5% between the A5 and B subtypes represented in the current study. The zoomed inset (shaded) highlights the range from 94% to 100%, to provide

clearer insight into less divergent gene regions. (B) Variations observed in the K4.2, ORF47 and K8 gene regions. Los Alamos Highlighter software was used to visualize mutations noted within the K4.2 (A), ORF47 (B) and K8 (C) gene regions. The plots show predicted amino acids changes as referenced in the figure legend, compared to the GK18 reference genome. Gaps in sequences are indicated by gray bars

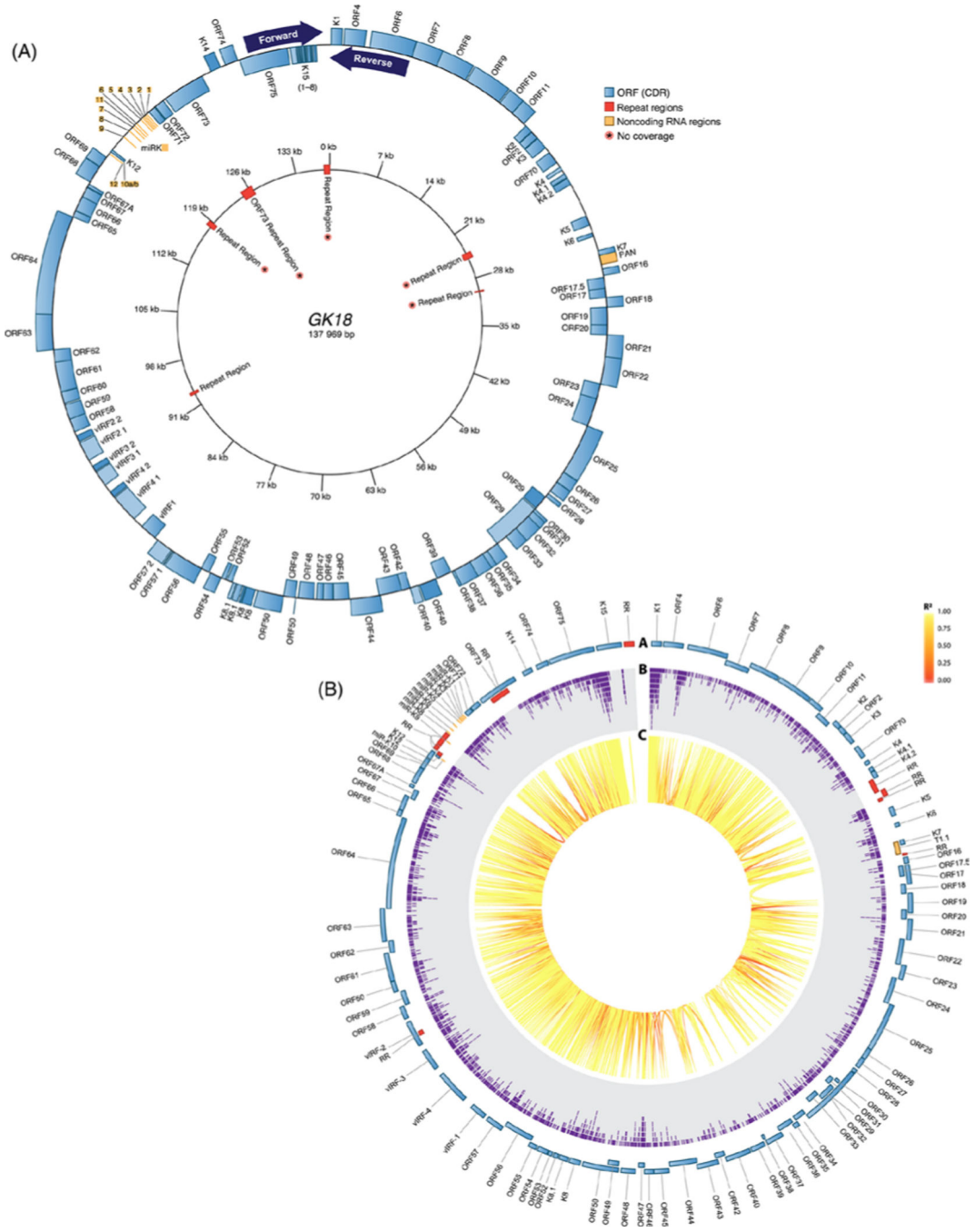


FIGURE 3.

Reference KSHV genome ORF (CDR) (GenBank. NC_009333). (A) K1 and K15 are indicated at the top center of the outer circle with the orientation of individual genes reflected in the placement on the outside or inside of the outer circle as shown. Repetitive regions, not sequenced by NGS method, are indicated by the inner circle. (B) Detailed mapping of the positions of single nucleotide variations, insertions and deletions (B ring) as well as linkage disequilibrium analysis (C ring) in the KSHV sequences analyzed from HIV+ participants. Masked repeat regions indicated in red. Graphs created with Circosplot

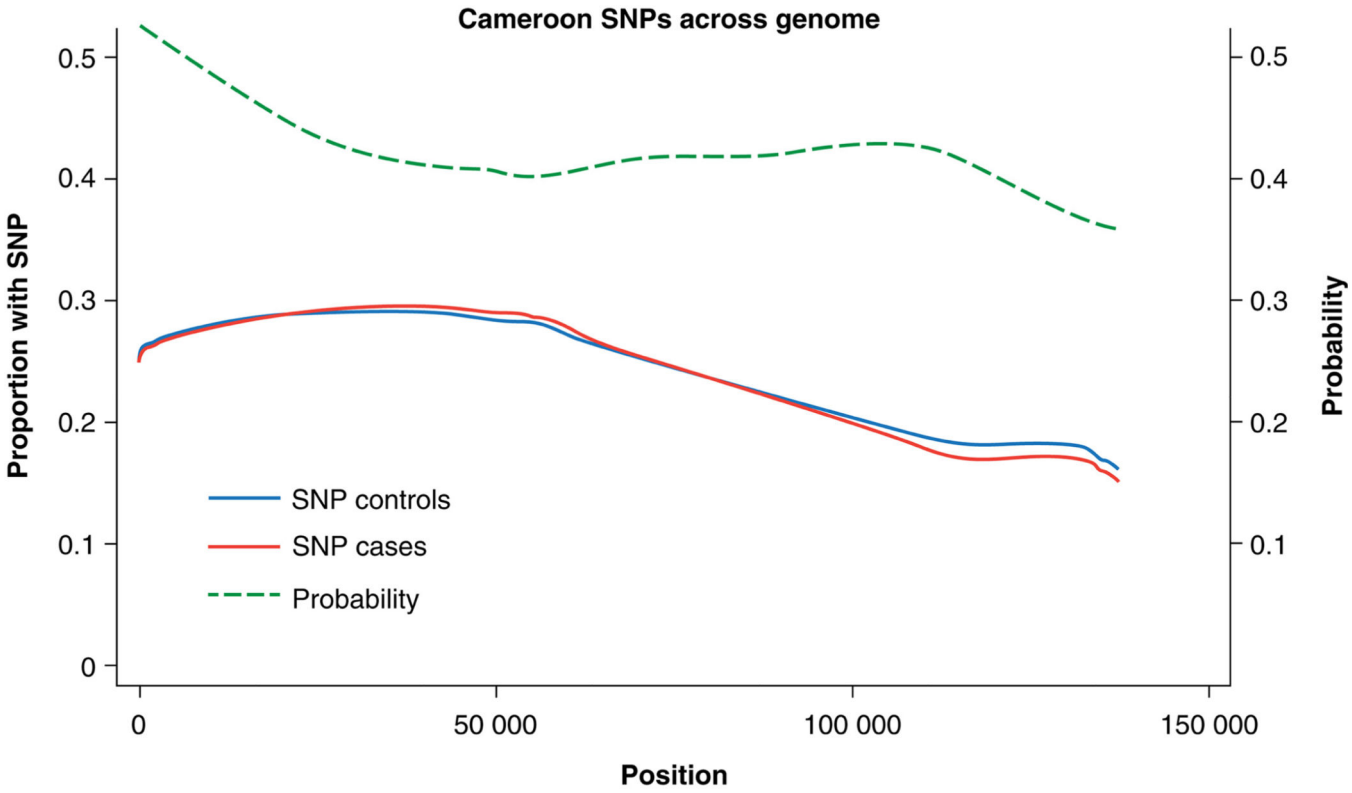
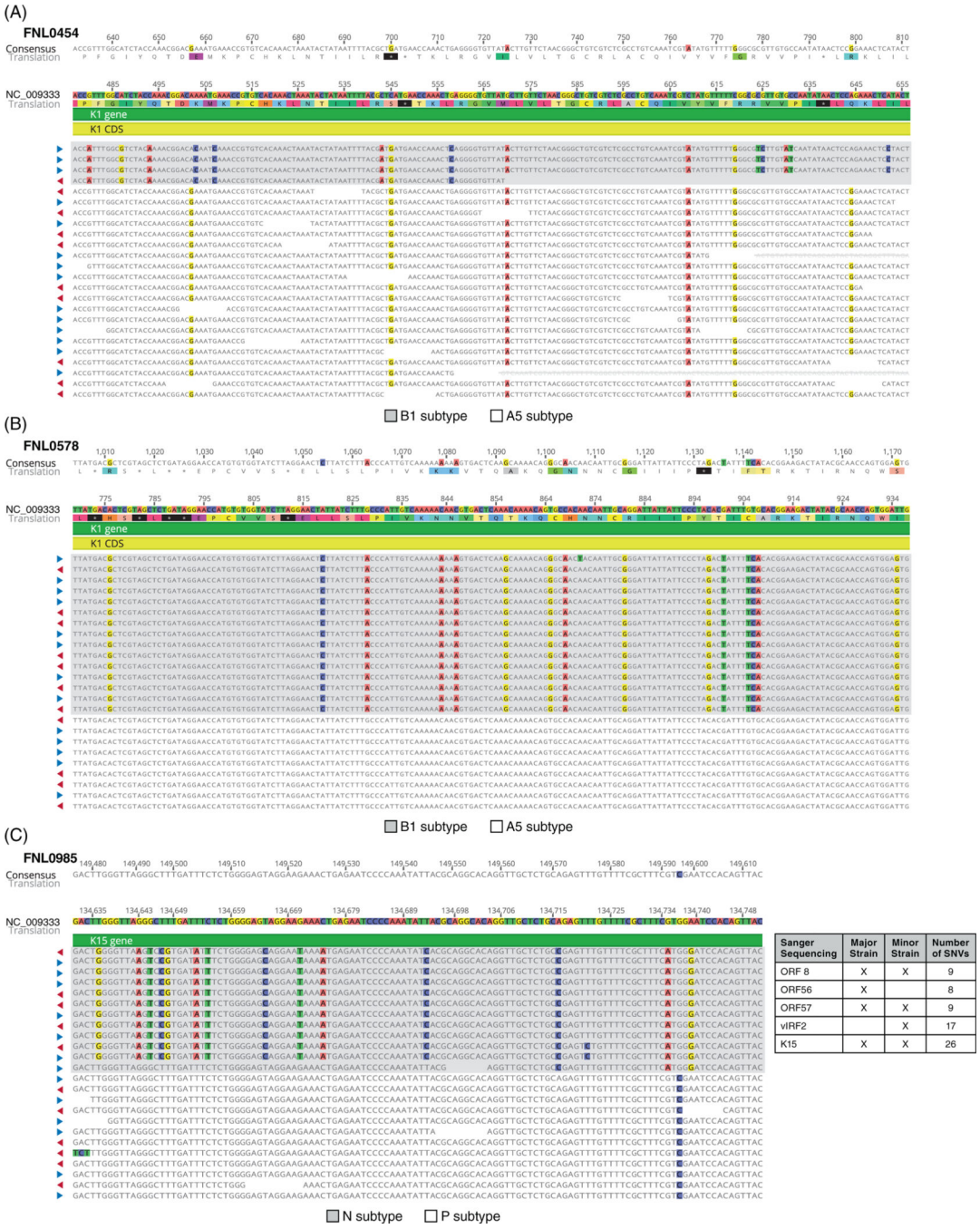


FIGURE 4.

KSHV sequence comparisons between AIDS-KS cases and controls. Proportion (left *y* axis) of KS (red) and control (blue) whole KSHV genomes with variant sequence compared to the reference GK18 at any given nucleotide position (*x* axis). A parametric test for the equality of proportion yielded the probability shown by the green line (right *y* axis) and did not identify any SNVs occurring more frequently in KS as compared to control sequences



Sanger Sequencing	Major Strain	Minor Strain	Number of SNVs
ORF 8	X	X	9
ORF56	X		8
ORF57	X	X	9
vIRF2		X	17
K15	X	X	26

FIGURE 5.

Mixed infection. KSHV sequence alignments are shown from the three identified dual infection samples. The alignments show a representative number of reads for the highly variable genes K1 (panels A and B) and K15 (panel C) using Geneious Prime software 2021.0.3. The shaded portion represents the A5 gene specific reads in the K1 gene alignments and the N subtype in the K15 alignment. For sample FNL0985, we were able to Sanger sequence confirm the major and minor KSHV strains in K15 gene, as well as

two additional gene regions attempted, as shown in adjacent table. Sanger sequence primer designs used to distinguish dual infections are included in Table S1

TABLE 1

Participant characteristics

ID	Material	Collection date	Ethnicity	HIV	Diagnosis	Sex	KSHVVL	K1/K15 subtype
FNL0075_CM	WBC	3/24/10	Bantu	HIV+	KS	M	35 000	B1/N
FNL0284_CM	WBC	7/11/11	Bantu	HIV+	KS	M	9120	A5/N
FNL0320_CM	WBC	6/18/10	Bantu	HIV+	KS	F	32 500	B1/P
FNL0346_CM	OF	6/22/10	Bantu	HIV−	KS	M	39 000	A5/P
FNL0417_CM	WBC	7/6/10	Semi-Bantu	HIV+	KS	M	55 000	A5/P
FNL0454_CM	WBC	7/22/10	Semi-Bantu	HIV+	KS	M	42 900	A5,B/P dual
FNL0456_CM	WBC	7/23/10	Bantu	HIV+	KS	F	27 700	B1/P
FNL0461_CM	WBC	7/27/10	Semi-Bantu	HIV+	KS	M	56 300	A5/M
FNL0467_CM	OF	8/25/10	Bantu	HIV+	KS	F	55 000	B1/P
FNL0515_CM	WBC/OF	8/31/10	Bantu	HIV+	KS	M	14 000/19 500	Deletion/P
FNL0526_CM	WBC	9/2/10	Semi-Bantu	HIV+	KS	F	13 200	B1/P
FNL0527_CM	OF	9/3/10	Hamites	HIV+	KS	M	7700	ND
FNL0531_CM	WBC	9/6/10	Semi-Bantu	HIV+	KS	F	32 600	A5/P
FNL0565_CM	WBC	9/15/10	Semi-Bantu	HIV+	KS	M	3000	ND
FNL0643_CM	OF	11/1/10	Semi-Bantu	HIV+	KS	F	24 000	B1/P
FNL0646_CM	WBC	11/2/10	Bantu	HIV+	KS	F	4860	ND
FNL0665_CM	WBC	11/3/10	Sudanese	HIV+	KS	M	38 500	A5/P
FNL0688_CM	OF	4/20/11	Bantu	HIV+	KS	M	26400	B1/N
FNL0696_CM	WBC	11/8/10	Hamites	HIV+	KS	M	4176	ND
FNL0749_CM	WBC	11/12/10	Bantu	HIV+	KS	F	1248	ND
FNL0814_CM	WBC	11/22/10	Semi-Bantu	HIV+	KS	F	18 900	A5/P
FNL0915_CM	WBC	12/8/10	Bantu	HIV+	KS	F	266400	A5/P
FNL0918_CM	WBC	12/9/10	Bantu	HIV+	KS	F	34 560	A5/P
FNL0972_CM	WBC	12/22/10	Bantu	HIV+	KS	F	8500	A5/P
FNL0985_CM	WBC	12/27/10	Bantu	HIV+	KS	F	20 450	A5/P,N dual
FNL0990_CM	WBC	12/29/10	Semi-Bantu	HIV−	KS	M	2520	ND
FNL1021_CM	WBC	1/12/11	Sudanese	HIV+	KS	M	3364	ND
FNL1034_CM	WBC	1/19/11	Semi-Bantu	HIV+	KS	F	71000	A5/M
FNL0057_CM	OF	3/19/10	Bantu	HIV+	Control	F	112 000	A5/P

ID	Material	Collection date	Ethnicity	HIV	Diagnosis	Sex	KSHVVL	K1/K15 subtype
FNL0058_CM	OF	3/19/10	Semi-Bantu	HIV+	Control	F	29 400	A5/P
FNL0087_CM	OF	3/24/10	Bantu	HIV+	Control	F	11 600	B1/P
FNL0163_CM	OF	4/14/10	Semi-Bantu	HIV+	Control	M	9200	A5/M
FNL0173_CM	OF	4/16/10	Semi-Bantu	HIV+	Control	M	10 500	B2/M
FNL0175_CM	OF	4/19/10	Semi-Bantu	HIV+	Control	M	12 600	A5/M
FNL0254_CM	WBC	12/10/10	Bantu	HIV+	Control	F	7050	B1/P
FNL0297_CM	OF	6/7/10	Bantu	HIV+	Control	M	36 000	A5/deletion
FNL0371_CM	OF	6/25/10	Bantu	HIV+	Control	M	8200	A5/P
FNL0395_CM	OF	6/29/10	Bantu	HIV+	Control	F	42 000	A5/P
FNL0432_CM	OF	7/14/10	Semi-Bantu	HIV−	Control	F	13 800	ND
FNL0498_CM	OF	8/25/10	Bantu	HIV+	Control	M	15 680	A5/P
FNL0536_CM	OF	9/7/10	Bantu	HIV+	Control	F	13 100	B1/N
FNL0537_CM	OF	9/7/10	Bantu	HIV+	Control	F	13 650	A5/M
FNL0578_CM	OF	9/21/10	Semi-Bantu	HIV+	Control	M	284 700	A5,B/P dual
FNL0651_CM	OF	11/2/10	Bantu	HIV+	Control	F	36100	A5/P
FNL0681_CM	OF	11/4/10	Bantu	HIV+	Control	M	41 160	ND
FNL0807_CM	OF	11/19/10	Bantu	HIV+	Control	M	72 200	A5/M
FNL0921_CM	OF	12/9/10	Hamites	HIV+	Control	M	18 900	B1/M
FNL0979_CM	OF	12/23/10	Bantu	HIV+	Control	F	6880	B1/P
FNL1020_CM	OF	11/12/11	Semi-Bantu	HIV+	Control	M	21 000	A5/P
FNL1040_CM	OF	1/20/11	Semi-Bantu	HIV+	Control	F	19 200	B1/P
FNL1106_CM	OF	2/21/11	Bantu	HIV−	Control	F	5000	A5/N
FNL1124_CM	OF	3/29/11	Bantu	HIV+	Control	M	18 600	B1/M

Note: Samples are listed referencing GenBank accession identifiers. Table includes self-reported demographics as well as the estimated KSHV genome copy number in the library preparations. K1/K15 subtype designations, as referenced in Figure 1, are included.

Abbreviations: OF, pelleted oral fluid; WB, whole blood extractions from Qiagen Paxgene.