

# Genome evolution of Kaposi sarcoma-associated herpesvirus (KSHV)

Razia Moorad,<sup>1</sup> Alice Peng,<sup>1</sup> Justin Landis,<sup>1</sup> Linda J. Pluta,<sup>1</sup> Patricio Cano,<sup>1</sup> Angelica Juarez,<sup>1</sup> Dirk P. Dittmer<sup>1</sup>

**AUTHOR AFFILIATION** See affiliation list on p. 18.

**ABSTRACT** Kaposi sarcoma (KS) is the most common cancer in people living with HIV (PLWH), particularly in sub-Saharan Africa (SSA), where Kaposi sarcoma herpesvirus (KSHV or human herpesvirus 8 [HHV-8]) is endemic. In KSHV endemic areas, the overall survival of KS patients has changed little over the past 20 years. A phylogenetic analysis of available full-length viral genomes ( $n = 164$ ) identified two different virus lineages that co-circulate in KSHV endemic regions today. Their sequences differ from the GenBank reference sequence and those of common laboratory strains, which originated in the 1990s in the US and Europe. Targeted short-read sequencing accuracy was validated by PacBio-based long-read sequencing to resolve repeats. This analysis identified over 1,000 single nucleotide variants (SNV) in a new 14-member sequence collection from tumor biopsies and blood in Malawi with  $127 \pm 32$  (median  $\pm$  SD) SNV per genome. Most were private, i.e., specific to one individual's virus. Within each of the two lineages, KSHV continues to evolve over time and across national borders by genetic drift and recombination. Analyses of shared SNVs by AlphaFold2 predicted some changes in the conformation of key viral proteins. These findings may help our understanding of herpesvirus evolution.

**IMPORTANCE** To understand viruses, the field needs to know their genetic makeup. To develop mechanistic models, targeted therapies, and vaccines, we need comprehensive and up-to-date sequence information on the viral strains that circulate where the diseases appear today. Our knowledge of Kaposi sarcoma herpesvirus (KSHV) sequence distribution and evolution is behind that of other human herpesviruses and RNA viruses. Here, we add to community knowledge using new technologies and artificial intelligence.

**KEYWORDS** Kaposi's sarcoma, human herpesviruses, KSHV, AlphaFold, PacBio, evolution, human herpesvirus 8, Kaposi sarcoma-associated herpesvirus, Kaposi sarcoma herpesvirus

Kaposi sarcoma herpesvirus (KSHV), formally known as Kaposi sarcoma-associated herpesvirus, is a double-stranded DNA virus. KSHV is also called human herpesvirus 8 (HHV-8). KSHV is the etiological agent for all forms of Kaposi sarcoma (KS) (1), for KS-associated diseases (KADs) (2, 3), the plasmablastic variant of multicentric Castleman's disease (MCD), and primary effusion lymphoma (PEL). KSHV also causes acute diseases such as KS-immune reconstitution inflammatory syndrome (KS-IRIS) and KSHV inflammatory cytokine syndrome (KICS) (reviewed in reference 4). There is no preventative vaccine against KSHV and no therapeutic vaccine against KS (reviewed in reference 5). KS is susceptible to broad-spectrum anti-herpesvirus agents such as ganciclovir, foscarnet, and cidofovir; however, these are not currently used to prevent or treat KS. There is no "cure" for KSHV or KS, as herpesviruses establish molecular latency for life.

**Editor** Lori Frappier, University of Toronto, Toronto, Ontario, Canada

Address correspondence to Dirk P. Dittmer, [dirkdittmer@me.com](mailto:dirkdittmer@me.com).

Razia Moorad and Alice Peng contributed equally to this article. The author order was determined by who initiated the project.

The authors declare no conflict of interest.

See the funding table on p. 19.

**Received** 5 November 2024

**Accepted** 11 February 2025

**Published** 16 April 2025

Copyright © 2025 Moorad et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

In sub-Saharan Africa (SSA), where KSHV is endemic and where the human immune deficiency virus (HIV) remains epidemic, KS is seen in children, adults, and the elderly with or without HIV co-infection (6). Here, KSHV is transmitted in infancy; seroprevalence can exceed 80% before sexual debut, and malaria is considered a cofactor for virus acquisition (7–11). KS is one of the most common cancers in the region (12).

With the widespread availability of combination antiretroviral therapy (cART), the 2000s witnessed a decline in the incidence rates of KS in men who have sex with men (MSM) populations (13). Today, however, KS incidence rates have stabilized. KSHV is firmly entrenched and in high-risk groups, seroprevalence approaches 40% (14). KS remains the most common cancer in people living with HIV (PLWH) worldwide. This stabilization of disease incidence, rather than continued decline, is due to several factors. First, KS incidence rates, independent of HIV, have always been age-dependent. Classic KS is a disease of the elderly in areas with KSHV prevalences upwards of 5%, such as in the Levant or 19th century Vienna. Organ-transplant-associated KS emerges independent of HIV (reviewed in reference 4). Lastly, as the initial cohort of PLWH ages, KS risk increases independent of HIV-related immune suppression. While cART alone slows the progression to AIDS and, by restoring CD4 counts, can cause KS regression in one-third of cases, other PLWH experience KS despite HIV suppression and near-normal CD4 counts (15–19). Today, “a high number of Kaposi sarcoma cases and disparities in Kaposi sarcoma outcomes persist in certain populations in the United States (20).” In 2020, KS was the most common cancer in men in Malawi, accounting for 16.7% of all new cancer cases and 5.7% in females, regardless of HIV status; in Uganda, KS incidence rates accounted for 17.1% of all new cancer cases in men and 7% of new cases in women (21).

As KS and KAD continue to determine overall mortality in PLWH, KSHV deserves continued attention, and KSHV sequence information requires continued validation. The question arises: how similar are the KSHV strains circulating in PLWH today to the original type specimen and the tool isolates that drive basic KSHV research, drug and vaccine development? The first step toward answering this question is a reexamination of the original KSHV isolates with third-generation long-read sequencing by synthesis methods.

The first complete KSHV genome, GenBank accession number [U75698](#), was derived from the Epstein-Barr virus (EBV)/KSHV-coinfected BC-1 PEL cell line in 1996 (22, 23): KSHV<sup>BC-1</sup>. The U75698 consensus sequence was obtained by Sanger-sequencing of a cosmid and lambda library. Next, the KSHV genomes from two primary KS lesions were obtained (U93872 and AF148805). These were also generated by Sanger-sequencing of cosmid or lambda libraries. Neither viruses nor cells corresponding to these two KS derived genomes were ever cultured or clonally expanded. The reference KSHV genome is strain KSHV<sup>GK18</sup>, based on GenBank entry [AF148805](#), represents overlapping cosmids cloned from a KS biopsy. The sequence is oriented as the unique (U) region followed by a single copy of the terminal repeat (TR). The mean number of terminal repeats and authentic ends of the linear virion genome were determined later and patched on. As noted in AF148805, “some copies of TR lack 39 bp” and “74 bp (were) deleted in cos19 and cos32, and also in the majority of DNA from the original biopsy; this disrupts the transmembrane anchor domain of the K8.1 protein; shown as wild type (undeleted).” Hence, it is possible that AF148805 was derived from a pool of defective genomes and repaired post hoc based on sequences from PEL cells. AF148805 became the current reference sequence, which is designated [NC\\_009333](#). No one has proven experimentally that the NC\_009333 sequence represents a replication-competent virus.

Two PEL cell lines deserve special attention: BCBL-1 and JSC-1. Both cell lines yield high levels of infectious KSHV virions. BCBL-1 only carries KSHV, and JSC-1 carries KSHV and EBV (24, 25). Replication-competent bacmids have been derived from these cell lines. These are BAC16 ([GQ994935](#), [MK143395](#), and [MK733609](#)) from the JSC-1 cell line (26) and BAC36 ([HQ404500](#)) and its derivatives from the BCBL-1 cell line (27). These bacmids are single clones of a viral episome and can be stably propagated in *Escherichia coli* (save for the terminal repeat region). The null hypothesis is that while PEL cell lines and KSHV genomes carried in them are under constant selection for faster growth, KSHV genomes

carried in bacteria with intact DNA repair and replication genes are only subject to neutral genetic drift.

According to the International Committee on Taxonomy of Viruses (ICTV), the family Herpesviridae belongs to the order Herpesvirales (28). The Herpesviridae includes three subfamilies: Alphaherpesvirinae, Betaherpesvirinae, and Gammaherpesvirinae. KSHV is of the genus Rhadinovirus, species Rhadinovirus humangamma8.

The community currently recognizes two types of Rhadinovirus humangamma8 based on the K15 sequences on the right-end side: the predominant (P) and minor (M) types (29, 30). The two forms of ORF-K15 have only 33% overall amino acid identity to one another but retain their conserved intracellular signaling motifs. The immediately adjacent ORF75 typically segregates with the M and P types of K15. Both K15 and ORF75 are leftward open reading frames (ORFs). A dual polyA site separates the leftward transcripts for K15 and ORF75 from the rightward K14/vGCPR transcripts. The historical designation as “type” is unfortunate since most evolutionary biologists use lineage as the next subdivision. Nevertheless, the designation “type” is used for KSHV and, more famously, human papillomavirus strains.

We had previously published the clinical data associated with a prospective cohort of HIV-positive KS patients in Malawi (31). Here, we describe the KSHV genomes from their tumor biopsies and plasma samples. We reported on the two African lineages of KSHV previously (32). Here, this finding is confirmed and extended in an independent cohort.

We provide findings of structural implications of shared non-synonymous single nucleotide variants (SNV), as modeled by AlphaFold. AlphaFold first performs a genetic database search based on the template sequence and second generates multiple sequence alignments (MSA), with regions of homology weighted by importance (attention). Third, a structured database search generates an MSA of aligned homologs per residue pair. These MSAs are imputed into the neural network block called the Evoformer, producing a  $N_{\text{seq}} \times N_{\text{res}}$  array ( $N_{\text{seq}}$  is the number of sequences;  $N_{\text{res}}$  is the number of residues) and a  $N_{\text{res}} \times N_{\text{res}}$  array, which represents the MSA and residue pairs, respectively (33). The Evoformer derives a structural hypothesis by exchanging information within the MSA and the residue pairs and extracts the spatial and evolutionary relationships of the protein template. The Evoformer then produces updated MSAs and residue pair representations, repeating the process several times before progressing to the structure module. Lastly, an AlphaFold2 structure is built in the structure module, followed by a local refinement to produce the final structure. While we are cautious about interpreting these models, the data predict that some African-lineage SNVs have functional consequences. Others change the virion surface and thus targets for antibody responses.

AlphaFold3 is now publicly available and free to use for any protein sequence. It can be used to explore the implications of structural variation. We explore three proteins as examples of how the new African Lineages of KSHV may alter virus protein structure. How these structural changes translate into functional phenotypes, however, is beyond even the best *in silico* predictions.

1. The ORF49 protein cooperates with ORF50 in activating several KSHV lytic promoters; additionally, ORF49 can activate the JNK/p38 pathways (34, 35). The crystal structure of ORF49 is solved ([5IPX.pdb](#); 2.8 Å) (36), providing a robust basis for AlphaFold2 predictions.
2. The ORF22, or glycoprotein H (gH), is one of eight glycoproteins associated with the viral envelope (37). The other envelope glycoproteins include gB (ORF8), gM (ORF39), gL (ORF47), gN (ORF53), ORF28, ORF68, and K8.1 (22). The gB, gH, and gL play crucial roles in viral attachment, fusion, and entry. These proteins are highly conserved across all herpesviruses, which aided in training AlphaFold. KSHV ORF22/gH has functional, sequence, and structural homologs in EBV, herpes simplex virus 1 (HSV-1), and human cytomegalovirus (HCMV), namely BXL2, UL22,

and UL75, respectively (22, 38). The gH and gL proteins form a non-covalently linked complex. The gL protein is required for the processing and intracellular transport of gH, indicative of a conserved, co-evolving interaction surface. The crystal structure of gH and gL in complex with the ligand binding domain (LBD) of EphA2 has been solved (39). There are three published X-ray structures for EBV BXLF2, two for HSV-1, one for pseudorabies virus, and three for HCMV UL75, which forms part of the pentameric entry complex. Neutralizing antibodies to the pentameric complex prevent HCMV infection (40, 41).

3. The ORF34 protein interacts with several other proteins, including ORF18, ORF23, ORF24, ORF31, and ORF66 (42). The interaction between ORF34, ORF31, and ORF23 is essential for late gene transcription: deletion of ORF34 results in the loss of progeny viruses. These proteins are highly conserved in EBV and HCMV (43) and bind together to form the viral pre-initiation complex (vPIC), which recruits the cellular RNA polymerase II. ORF34 acts as a scaffold for the vPIC proteins. The central region of ORF34 binds cellular transcription-associated factors, whereas the C-terminus of ORF34 recruits ORF24. ORF24 is a TATA box-binding protein homolog that binds to the TATT sequence in the late gene promoters of KSHV (42–44).

How does one assess the quality of the structure prediction? AlphaFold2 provides multiple scoring metrics, such as the predicted local distance difference test (pLDDT), representing the predicted side-chain torsional angles and accuracy per structure residue (33). The scoring ranges from 0 to 100, where regions of a pLDDT >90 are considered highly accurate. Regions with a pLDDT between 70 and 90 provide a good backbone structure, and regions with a pLDDT between 50 and 70 are modeled with low confidence. Regions with a pLDDT below 50 have low confidence and should not be interpreted.

Another confidence score is the predicted aligned error (PAE). This has two meanings. If there exists a structure, also called a ground truth, PAE measures how well a prediction matches the true structure for each amino acid position in three-dimensional space. This changes for each amino acid position within a protein. Since AlphaFold is based on sequence alignments, more and more closely related sequences will have higher PAE scores. Whereas, PAE is a prediction, root mean square deviation (RMSD) is measured between exactly two final structures. If there is no structure to measure the AlphaFold prediction against, PAE is a computational estimate of how confident the model is about the prediction. In Fig. 5C and 6B, the color at position x, y indicates the AlphaFold2 expected position error at residue x if the predicted and actual structure were aligned at residue y. A low PAE for residue pairs x and y from two different domains indicates that AlphaFold2 had confidently predicted their relative position and orientation. Conversely, a high PAE for residue pairs x and y would indicate low confidence in their predicted positions and orientation. Below, both a structural model and error estimators are reported.

## MATERIALS AND METHODS

### Sample collection

Tissue biopsies from KS skin lesions and plasma samples were collected from 137 participants (31). PBMCs were prepared as cell blocks by centrifugation, with the addition of a thrombin solution for cryopreservation and biopsies embedded in formalin-fixed paraffin-embedded (FFPE) blocks.

### DNA isolation

DNA was extracted from FFPE curls using the QIAmp DNA FFPE Tissue Kit (Qiagen, #56404). DNA was extracted from plasma samples using the Roche MagNA Pure

Compact Instrument (Roche, #0373114601). Ten microliters 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. The quantity and quality of the DNA and RNA were assessed by Nanodrop (Thermo Fisher Scientific), Qubit (Thermo Fisher Scientific), and TapeStation (Agilent Technologies).

### Short-read sequencing

DNA library preparation and sequencing from plasma samples were performed using the automated approach on the Ion Torrent Genexus Integrated Sequencer (Thermo Fisher Scientific; Pub. No. MAN0017912), using custom primers to enrich for KSHV. For FFPE samples, library preparation and sequencing were automated on the Ion Torrent Ion Chef and S5 Sequencer with default parameters using custom primers (Thermo Fisher Scientific). Barcodes, primers, and adapter sequences are removed before FASTQ sequence data files are output from the Ion Torrent S5 and Genexus sequencers, respectively. Previously reported primers and PCR conditions (32) were used to amplify and Sanger sequence the GC-rich and repetitive regions not captured by targeted short-read sequencing.

### Long-read sequencing

Approximately 2.5  $\mu$ g of genomic DNA was fragmented using G-Tube (Covaris) following the technical note "Covaris g-Tube DNA Shearing for SMRTbell Prep Kit 3.0." Fragmented DNA was then purified using 1 $\times$  SMRTbell cleanup beads. The size was determined by TapeStation. Concentration was determined by Qubit. The Hi-Fi library was made using the SMRTbell Prep kit 3.0 (procedure and checklist 102-166-600 REV02 MAR2023). Sequencing was conducted per the manufacturer's instructions (SMRTLink user guide—102-278-200 Version 01 [April 2022]) and SMRT sequencers operations guide (101-774-06 Pacific Biosciences of California, Inc. ["PacBio"]).

### Quality trimming, filtering, and read mapping

FASTQ reads were imported into CLC Genomics Workbench v23.0.2 (Qiagen Inc.). If the quality control report indicated a specific number of bases of poor quality, this number was used for the trim function. If the quality control report did not yield a particular, fixed number of bases to trim, reads were trimmed by removing low-quality base pairs (quality limit = 0.05), including 40 nucleotides at the 5' terminal and 10 nucleotides at the 3' terminal. All sequences shorter than 50 nucleotides were filtered out. Next, reads were mapped to the KSHV genome (NC\_009333) using the CLC Genomics Workbench v23.0.2 (Qiagen) map-to-reference tool with the following parameters: no reference masking and default mapping options (match score = 1, mismatch score = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.5, similarity fraction = 0.8). The mapped reads were trimmed to the reference. Non-specific matches were ignored. Duplicate mapped reads were removed using default parameters on the remove-duplicate-mapped-reads tool on CLC Genomics Workbench v23.0.2. The CLC Extract Consensus Sequence tool was used to create a consensus sequence. The low coverage threshold was set to three (this was adjusted based on the low coverage areas and the number of reads mapped there in the map file). Conflicts were resolved by voting using quality scores. In the GenBank records, low coverage regions are indicated with "N." For the phylogenetic analysis of proteins and structure prediction, the open reading frames (ORF) in the consensus sequence were repaired, and SNVs resulting in internal stop codons were corrected. The final, repaired, and annotated sequences were used for further analysis and deposited in GenBank (Table 1).

### Variant calling

Variant calling was performed using two orthogonal methods. First, using raw reads, variants, including single nucleotide variants (SNVs), multiple nucleotide variants (MNVs),

TABLE 1 KSHV whole genome sequence metrics<sup>a</sup>

GenBank accession no.	KSP	Type	Age	Gender	No. of sequencing reads	No. of reads after trim	No. of mapped reads	Duplicate mapped reads	Avg. read length	Fold coverage
PQ899225	KSP009	Plasma	33	Male	1,538,048	264,000	251,720	179,991	138	457
PV032230	KSP015	Plasma	49	Male	1,344,479	223,940	210,959	59,578	140	214
PV032229	KSP012	Plasma	41	Male	1,323,840	111,788	108,407	46,496	150	118
PV061047	KSP090	Plasma	42	Male	1,213,127	109,667	91,928	35,438	131	87
PV016613	KSP002*	Plasma	49	Male	1,410,476	395,816	373,504	107,524	139	376
PV032228	KSP010	Plasma	37	Male	1,484,612	148,002	122,055	48,903	146	129
PV032227	KSP008	Plasma	35	Male	1,484,539	230,510	212,436	66,490	144	221
PV016612	KSP001	Plasma	42	Female	1,318,410	195,553	186,986	112,975	146	198
PV061044	KSP024	FFPE	37	Male	572,368	175,078	130,953	89,215	136	129
PV061045	KSP022	FFPE	41	Male	940,920	436,769	371,591	266,280	150	403
PV188213	KSP017	FFPE	37	Male	27,660,787	24,309,935	23,809,123	4,021,265	142	24,468
PV016611	KSP021	FFPE	28	Male	23,723,381	20,459,107	18,268,644	15,213,321	140	18,507
PV061046	KSP031	FFPE	53	Male	24,656,689	31,302,400	30,784,848	4,089,032	126	28,078
PQ899226	KSP002*	FFPE	49	Male	3,497,329	2,851,552	2,090,124	1,655,918	139	2,103

<sup>a</sup>The metadata and mapping results of the sequenced KS samples from Malawi. As indicated, samples were isolated from plasma and KS lesions in FFPE blocks. The sequences were submitted to GenBank, and their accession numbers can be accessed as described in Materials and Methods. \* indicates two samples from the same person.

and insertions and deletions (INDELs), were called using the Basic Variant Detection tool on CLC Genomics Workbench v23.0.2 (Qiagen), with default parameters. SNVs with a minimum average Phred quality score >20, a minimum frequency >35%, and a minimum coverage of >40 reads were identified. A base quality filter with a minimum central quality of 20 and a minimum neighborhood quality of 15 was selected. Pyro-error variants were removed in homopolymer regions of a minimum length of three nucleotides. Second, the SNV calls variant calling on the MAFFT alignment of finished consensus sequences was performed using the find variations/SNPs tool implemented in Geneious v11.1.5 (Dotmatics, Inc.).

## Phylogenetic analysis

The phylogenetic tree was constructed using the high-coverage KSHV genomes generated in this study and other previously published KSHV genomes isolated from Africa. These include those isolated from Zambia (45), Uganda (46, 47), Malawi (32), and Cameroon (48). Additional publicly available genomes isolated from the USA and Japan were included in the analysis along with the reference genome [NC\\_009333](#) (49).

The KSHV sequences used herein were retrieved using “human gammaherpesvirus 8,” which is the official GenBank taxonomy of KSHV (NCBI Taxonomy ID 37296) across all fields AND filtering by sequence length greater than 100,000. The search result changes with calendar time. For batch download by accession number, the NCBI ncbi-entrez-direct command was used. Note that the GenBank naming rules do not follow the naming conventions of the ICTV (<https://ictv.global>). The ICTV 2022 taxonomy for KSHV is Duplodnaviria, Heunggongvirae, Peploviricota, Herviviricetes, Herpesvirales, Orthoherpesviridae, Gammaherpesvirinae, Rhadinovirus, and Rhadinovirus humangamma8.

All genomes were aligned using MAFFT, implemented in Geneious v9.1.8 with default parameters (50). The conserved long unique region (LUR) from ORF16 to ORF58 was extracted from the multiple alignments of 164 genomes, resulting in 65,840 bp of continuous sequence, used to infer the Bayesian Maximum Likelihood (ML) tree implemented in Beast v1.10.4 (51, 52). The HKY substitution model, with estimated base frequencies, and the gamma site heterogeneity model were selected to estimate site substitution rates. A strict clock and a constant size coalescent restriction were chosen to construct the ML tree. The ML tree was rooted in the KSHV reference genome [NC\\_009333](#). The phylogenetic tree representing the left and right side of the LUR (total length is 33,299 bp and 49,517 bp, respectively) was extracted from MAFFT whole



**TABLE 2** Variants identified in the KSHV LUR<sup>a</sup>

Count	Position	Reference	SNV	ORF	Amino acid	Lineage
1	43962	G	T	ORF25	Ala363Ser	A
2	48197	G	C	ORF27	Leu75Phe	A
3	48308	G	T	ORF27	Glu112Asp	A
4	54964	T	A	ORF34	Leu64Gln	A
5	55059	G	A	ORF34	Glu96Lys	A
6	59464	C	T	ORF39	Val271Ile	A
7	61165	T	A	ORF40	Ser253Arg	A
8	61698	A	G	ORF40	Lys431Arg	A
9	70640	C	A	ORF48	Val281Lue	A
10	75091	T	G	K8	Val48Gly	A
11	76356	G	A	K8.1	Val115Met	A
12	78886	T	A	ORF55	Glu221Val	A
13	81515	G	A	ORF56	Ala661Thr	A
14	81531	C	T	ORF56	Ala666Val	A
15	82616	A	G	ORF57	Asn114Asp	A
16	88416	G	A	vIRF-4	Val165Ala	A
17	88431	T	C	vIRF-4	Tyr160Cys	A
18	88438	G	C	vIRF-4	Gln158Glu	A
19	93230	A	T	vIRF-2	Phe293Lue	A
20	93346	C	T	vIRF-2	Val255Met	A
1	38427	G	A	orf22	Asp406Asn	A&B
2	55446	G	T	ORF34	Ala225Ser	A&B
3	63357	C	T	ORF42-43	Gly566Ser	A&B
4	70300	C	A	ORF48	Cys394Phe	A&B
5	70792	C	T	ORF48	Ser230Asn	A&B
6	92640	A	G	vIRF-2	Leu490Ser	A&B
1	30759	G	C	ORF16	Ser173Thr	B
2	52693	T	C	ORF32	Ser397Leu	B
3	54954	T	A	ORF34	Cys61Ser	B
4	92107	T	C	vIRF-2	Ile668Val	B
5	92124	C	G	vIRF-2	Trp662Ser	B

<sup>a</sup>The 31 variants identified in the LUR of the Malawian samples are described in this study. In subtype A, 20 variants were identified, five in subtype B, and six variants were shared between subtypes A and B. Position numbers are based on the reference genome (NC\_009333).

genome alignment and constructed using RAxML, implemented in Geneious v.9.1.8, with default parameters (53). The ML trees were constructed using the same parameters described above.

The Splits Tree analysis was performed to investigate conflicting phylogenetic signals and was performed on the ML tree alignment of the LUR using SplitsTree4 (54). The neighbor-net split network was constructed using uncorrected P character transformation and excluded gap sites. Statistical analysis included performing 1,000 bootstrap replicates and the Phi test for evidence of recombination.

## AlphaFold analysis

AlphaFold2 (AF2), V2.3.1, was downloaded from the Google DeepMind GIT repository and installed on a high-performance computing server. UCSF ChimeraX v1.6.1 was used to generate models of the same ORF proteins. KSHV ORF proteins with a non-synonymous variant (Table 2) were modeled using protein sequences from a Malawian isolate in the African subtypes A (MZ712180) and B (MZ712182) identified in the phylogenetic tree, including the reference genome NC\_009333. Default monomer-PTM parameters were selected with the option to use PDB structures in the alignment and energy minimizations to optimize the structure. The PDB files of the best-ranking protein structure

models were viewed using Pymol v2.5.4. Updated predictions can be obtained using the public AlphaFold3 server at <https://alphafoldserver.com>.

## RESULTS

### Validated genomes from the pre-NGS era increase rigor, robustness, and reproducibility

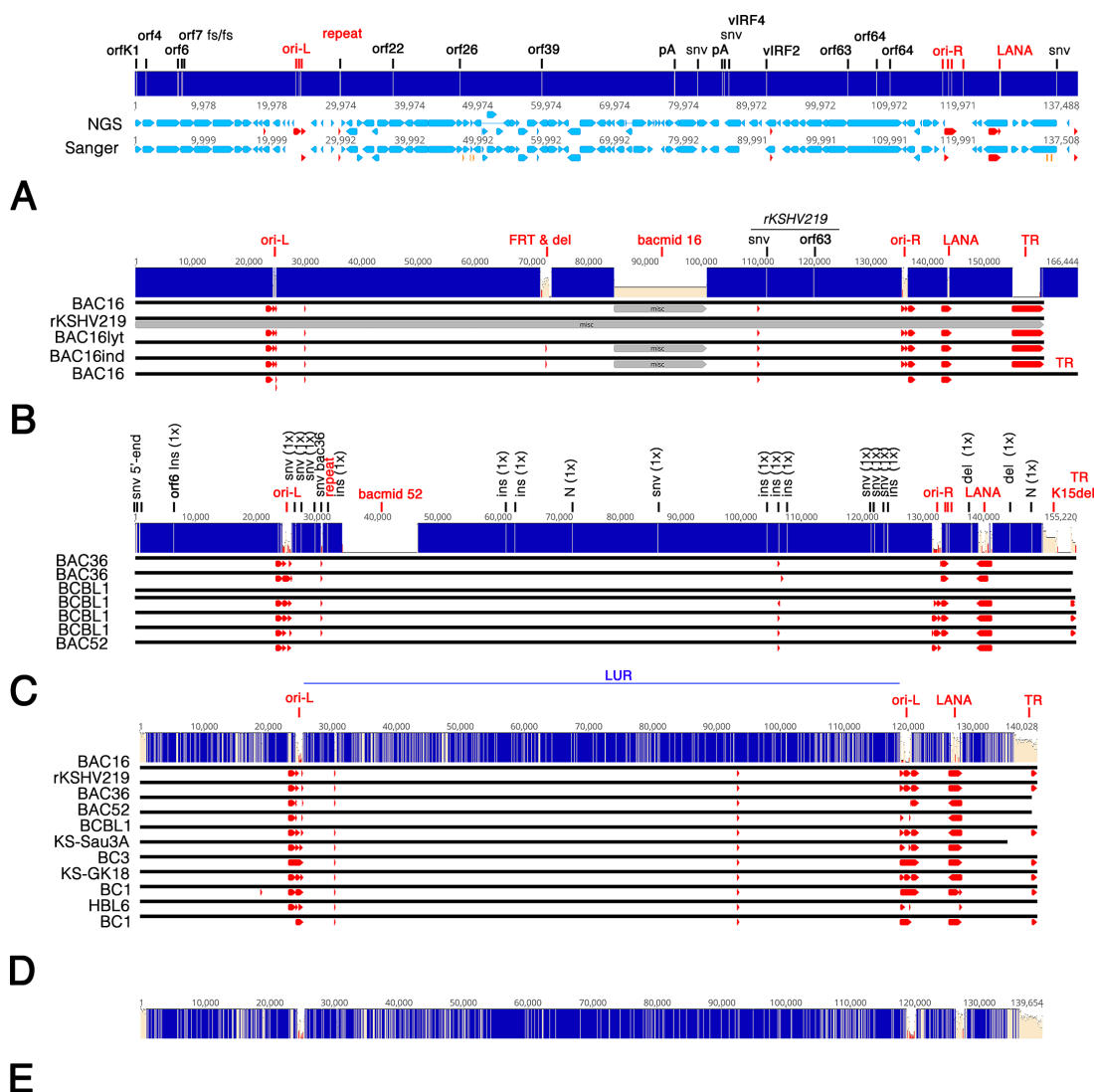
Most KSHV whole genome sequences in GenBank were derived by short-read next-generation sequencing (NGS), also called sequencing-by-synthesis. These include multiple, independent entire genome sequences from the same PEL cell lines. Figure 1A compares the KSHV<sup>BC-1</sup> consensus derived targeted by short-read NGS ([MK733607](#)) in 2019 to the Sanger-sequencing derived sequence for KSHV<sup>BC-1</sup> (NC\_003409, U75698) published in 1996 (22). SNVs outside the repeat regions appeared at random. All ORFs stayed “in-frame,” resulting in single-nucleotide and amino acid SNVs. None yield truncated proteins. There were  $n = 50$  (0.05%, equivalent to a PHRED value of Q40) nucleotide changes between the original Sanger sequence (U75698) and the short-read NGS, not counting variation in the repeat regions. Whether these are sequence errors in the short-read sequencing, the original GenBank entry, or whether the BC1 episomes accumulated mutations while in continuous culture is unknown.

Only the intended mutations and repeat regions differed among the different BAC16 genomes derived from the PEL cell lines JSC-1 by short-read sequencing (55) (Fig. 1B). Outside the repeat regions, all bacterially-propagated, single-origin viral genomes differed by two SNVs from the rKSHV219.1 maintained in mammalian cells and created in the mammalian JSC-1 cell line “by transfecting JSC-1 cells with pQ219, a plasmid which contains a 4.8 kb segment of KSHV DNA (from BC1 cells) with the RFP/GFP/PURO cassette inserted between ORF K9 and ORF 57 at a site that sequence data indicates does not contain a gene” (56). Based on this alignment, we estimated the minimal error due to neutral genetic drift in bacteria for high-coverage, high-confidence sequences at 2/160,465 or 0.0012%. The bac-propagated viral genomes were 99.9998% identical. This result instills confidence in KSHV bacmid-derived viral mutants and lays to rest concerns that KSHV bacmids are unstable in *E. coli*. There was no evidence of selection against toxic proteins or genome fragments in *E. coli*, except the TR.

Among the different published BAC36 and BCBL-1 genomes, the changes were single SNVs and short INDELs in one or another record but not shared across multiple GenBank entries. These were likely due to sequencing errors. The alignment confirmed the designed mutations, such as a deletion of K15 (57), insertion of the BAC52 backbone (58), and variation in the repeat regions (Fig. 1C). The BAC36 clone has a duplicated unique region fragment inserted in the TR (59). At the time of writing, the following KSHV strains are available as a genome sequence in GenBank and as biological material in public repositories: KSHV<sup>BC-1</sup>, KSHV<sup>BCBL-1</sup>, KSHV<sup>JSC-1</sup>, and KSHV<sup>HBL6</sup>. Of those, the KSHV<sup>BCBL-1</sup> and KSHV<sup>JSC-1</sup> were experimentally proven as infectious and replication-competent.

Variation was much more considerable among a representative collection of pre-2000 PEL and KS-derived KSHV genomes (Fig. 1D). The median difference for 55 pairwise comparisons was 1,977 bp, with a mean  $\pm$  SD of  $2,212 \pm 1,131$  bp. Much of the difference resides in the ends, as different sequence entries include more or less nucleotides at the extreme 5' end (approximately  $\pm 20$  bp) and the 3' end. To account for the possibility that KS isolates introduce extra variation, the multiple alignment was repeated with just the nine PEL isolates (Fig. 1E). Even among the KSHV PEL isolates collected in a similar data range, variation was noticeable. The medium pairwise similarity was 98%, with a mean  $\pm$  SD of  $98\% \pm 0.88\%$ . Note that 1% of 139,654 bp is still 1,396 individual SNVs. These nucleotide sequences form the historical sequence space and framework in which current experimental work is conducted.





**FIG 1** Overview of KSHV sequences. (A) Comparison of BC1 isolates U75698 obtained by Sanger-sequencing (Sanger) of a cosmid library and MK7333609 obtained by genome-enriching NGS (NGS). MK7333609 is missing the first 20 nucleotides present in U75698. The ORFs are shown in blue and repeat sequences in red. Individual changes are indicated as white lines in the identity bar and annotated above. Regions of 100% identity are in blue. (B) Comparison of non-identical bacmid BAC16 whole genome sequences (GQ99435, KY246443, KY246444, MK208823) and the rK219 isolate (KF588566). Entry OK358814 was 100% identical to GQ994935. Individual changes are indicated in the identity bar and annotated above. (C) Comparison of non-identical bacmid BAC36 and BCBL-1 whole genome sequences HQ404500, JX228174, MN205539, MT936340, MZ712172, OR117738, OR333977. (D) Comparison of early PEL, KS, and bacmid-contained KSHV genomes: JSC-1<sup>BAC16</sup> (GQ994935), JSC-1<sup>rKSHV219.1</sup> (KF588566), BC-1<sup>Russo</sup> (U75698), BC-1<sup>unc</sup> (MK7333607), BC3 (MK87631), BCBL-1<sup>PacBio</sup> (OR117738), BCBL-1<sup>BAC36</sup> (HQ404500), BCBL-1<sup>BAC52</sup> (OR333977), KS<sup>GK18</sup> (NC\_009333), KS<sup>Sau3a</sup> (U93872), HBL6 (OR573937). Individual changes are indicated in the identity bar. Also shown is the LUR (from ORF16 to ORF58) between the two Ori-Lyt sites comprising 65,840 bp of continuous sequence. (E) Identity graph for just the PEL isolates of (D).

### Additional, full-length viral genomes from PEL cell lines

DNA from the BCBL-1, JSC-1, VG-1, BC3, and BC-1 cell lines was subjected to high-fidelity (PacBio HiFi), long-read sequencing. This approach yielded complete, repeat-corrected KSHV<sup>BCBL-1</sup> (OR117738), KSHV<sup>BC3</sup> (PQ787490), KSHV<sup>VG-1</sup> (PQ899224), KSHV<sup>JSC-1</sup> (PQ868890), and KSHV<sup>BC-1</sup> (PQ868889) genomes (Fig. 2A through C) and the EBV<sup>BC-1</sup> and EBV<sup>JSC-1</sup> genomes (Fig. 2D through F). For this experiment, the viruses were not reactivated, and viral DNA was not enriched; instead, total high molecular weight DNA was used, assuming that the KSHV plasmids were present in multiple copies per cell (60–63). In the

circular plasmids, the TRs are continuous and linked to unique sequences on either end. Average coverage was 20x for the human genome and higher for the multi-copy KSHV plasmids (Fig. 2F). We compared two approaches to genome assembly. First, we mapped reads directly to the reference genome. Coverage by long-read NGS was lower than by short-read NGS, but PacBio HiFi sequencing approximates Q50 base-pair quality, which exceeded that of Sanger sequencing (Q20). Second, we *de novo* assembled the viral genomes from human-sequence-depleted reads. Long-read NGS yielded a single contig. *De novo* assembly more accurately captures insertions or sequences not previously reported. There were none.

### Additional full-length viral genomes from KSHV endemic regions

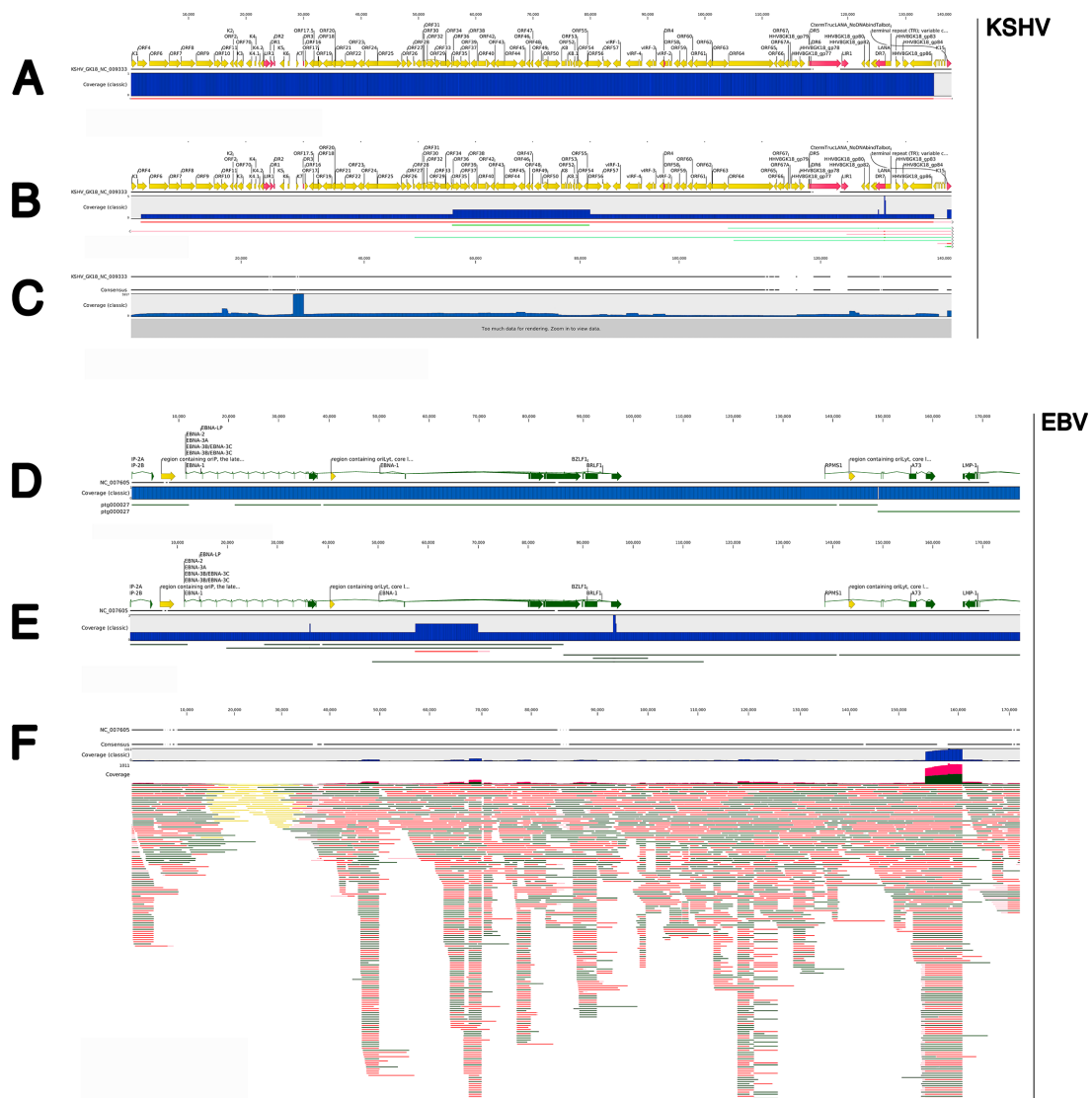
Phylogenetic and structural predictions are only possible due to publicly shared data. We and others continue to determine and share high-quality, full-length KSHV genomes through GenBank. Due to human subject protection regulations, many entries contain only minimal identifying information. Here, we report additional pre-cancer, pre-cART treatment KSHV genomes from HIV-positive participants (31) using targeted amplification (Table 1). All primer sequences were removed before mapping. The average coverage for the KSHV genomes isolated from FFPE biopsies was 12,206-fold. The coverage for KSHV genomes derived from KS biopsies was significantly higher than the average genome coverage isolated from plasma, which was 199-fold. This difference was consistent with KSHV biology. In the absence of concurrent KAD, the KSHV DNA copy number in the plasma of HIV-positive KS patients is low compared to other herpesvirus infections (6, 31, 32, 64, 65). Compared to plasma, DNA copy number is higher in the saliva (47, 66–68). For benchmarking, note that the genome quality of early GenBank entries, and presumably the KSHV reference genome, had a minimum coverage of 12-fold (22).

Calling KSHV genomes full-length based on short-read NGS alone is a misnomer because short reads cannot resolve longer repeat regions. These repeat regions include, most notoriously, the terminal (TR) and LANA central domain repeats (69). These pose problems in the initial targeted enrichment, sequencing, and alignment. For clinical samples, separate PCR amplification followed by Sanger sequencing of these regions was conducted (32). Because hypervariable regions are also hyper-recombinogenic, two different batches of the same cell line may have differing numbers of these repeats. The GenBank entry represents the most prevalent variant for repeat regions (TR, ori-L, ori-R, LANA, K15). One way to overcome the conundrum of herpesvirus repeats was to deploy long-read sequencing.

### Phylogenetic reconstruction of KSHV circulating in KS endemic regions

To reconstruct the evolution and population structure of KSHV lineages circulating today,  $n = 164$  full-length genomes isolated from Africa were used to construct the Bayesian ML phylogenetic tree. The tree was constructed using an alignment of the central conserved LUR of KSHV (from ORF16 to ORF58) between the two Ori-Lyt sites comprising 65,840 bp of continuous sequence. The LUR excluded the hypervariable K1 and K15 genes, which are subject to strong selection and less helpful in discerning long-term evolutionary patterns. Put differently, nucleotide substitutions within the LUR likely represent genetic drift driven by random single nucleotide mutations. In contrast, SNVs outside the LUR are closely linked to K1 and K15 genotypes. These two genes are under heavy immune selection. This selection may include convergent evolution and recombination, which pose difficulties for phylogenetic topology reconstruction.

Building a phylogeny on just the LUR tests the hypothesis that there is continuing drift of KSHV in endemic regions, where the virus is primarily transmitted from mother to child during the early years of infancy (8, 70–72). The transmission pattern in SSA differs from the MSM cohort of PLWH in the US, which may have started with a recent, single seeding event (73). As a result of different patterns of evolution, functional differences in



**FIG 2** Alignment summary of untargeted, PacBio-derived KSHV and EBV genomes for BC-1 cells. (A)  $BC_1$ KSHV<sup>PacBio</sup> *de novo* assembled from human-depleted reads. The lines diagrammed below indicate *de novo* assembled contigs in either the forward (green) or reverse (red) direction. (B)  $BC_1$ KSHV<sup>PacBio</sup> *de novo* assembled from all reads. The lines diagrammed below indicate *de novo* assembled contigs in either the forward (green) or reverse (red) direction. (C)  $BC_1$ KSHV<sup>PacBio</sup> consensus based on long-read sequencing mapped to GenBank reference NC\_009333, a K15 type P strain, vs BC1, which is the M strain. (D)  $BC_1$ EBV<sup>PacBio</sup> *de novo* assembled from human-depleted reads. The ptg00027 designation refers to *de novo* assembled contigs, which are diagrammed as lines below. (E) *De novo* assembled from all reads, with the black lines below indicating the longest contigs originating from the assembly. (F) Individual PacBio long-reads mapped to GenBank reference NC\_007605. Shown in yellow and green are selected ORF and transcripts. Shown in blue are the coverage graphs (solid blue indicates a single, the longest, contig aligned). Individual reads are green (forward) or red (reverse). Reads in yellow are non-unique mappings.

infectivity, reactivation from latency, replication, and transmissibility may have evolved differently and now define KS in SSA.

When the tree was rooted in the European KSHV reference genome NC\_009333 (GK18), the topology confirmed the existence of the two subtypes of KSHV within SSA (32) (Fig. 3A). All European or USA genomes clustered together and apart from the SSA genomes. Each SSA subtype contains genomes from the East African countries of Zambia (45), Uganda (46, 47), and Malawi (32), suggesting a concurrent circulation of both subtypes. Since the different studies used different sequencing technologies, the two subtypes are not the result of method biases. Genomes isolated from the West African Nation of Cameroon (48) fell within one sublineage. Except for some genomes from Uganda, the SSA KSHV genomes did not intermix with the European ones. The

KSHV genomes reported from Japan clustered apart from European or African isolates (Fig. 3, asterisk) (74). The genomes obtained in Malawi, as described in this study and previously, were characterized as either lineage A or B based on the tree of the LUR. We identified 20 non-synonymous variants in sublineage A and five in sublineage B. An additional six non-synonymous variants were shared across all lineages (Table 2).

To extend our analysis beyond the central region, we explored alignments of regions to the left and right of the central LUR. The left side harbors the hypervariable K1 gene and the left origin of replication (Fig. 3B). The hypervariable region within the K1 ORF drove the phylogeny, as genomes clustered by their K1 subtype. The new genomes reported here were of the A5 subtype, typical for African isolates (75–77) and separated from the European and typically older strains of the C3 subtype. The phylogeny to the right of the LUR was driven by the hypervariable K15 gene (Fig. 3C), as genomes clustered by their K15 subtype (30, 78, 79). The new genomes described in this study are of the P subtype. Based on K15 and adjacent regions of the viral genome, the African isolates clustered apart from the early US and European isolates.

A SplitTree network analysis (80) on the LUR was conducted to formally investigate recombination events. In contrast to the ML trees above, this analysis was unrooted. The SplitTree network algorithm also divided the genomes into three partitions consistent with the rooted ML tree. The darker shading of the branches in the SplitTree network corresponds to the higher confidence of the branch. The SplitTree analysis supports the phylogenetic clustering of the genomes into three sublineages, two consisting of recent genomes from SSA and one containing the genomes from Western Europe and the US (Fig. 4). Within this network, a Phi test indicated statistically significant evidence for recombination ( $P < 0.01$ ). These findings substantiate the genetic variability found within the LUR region of KSHV and provide evidence for recombination events within the evolutionary history of KSHV.

In conclusion, KSHV seems to drift over time and space, with more recent genomes from KSHV endemic regions separating into two lineages and differing from calendar-date-older sequences collected from US and European isolates. This conclusion, however, is confounded by the fact that sequencing technologies and bioinformatics assembly methods also developed over calendar time.

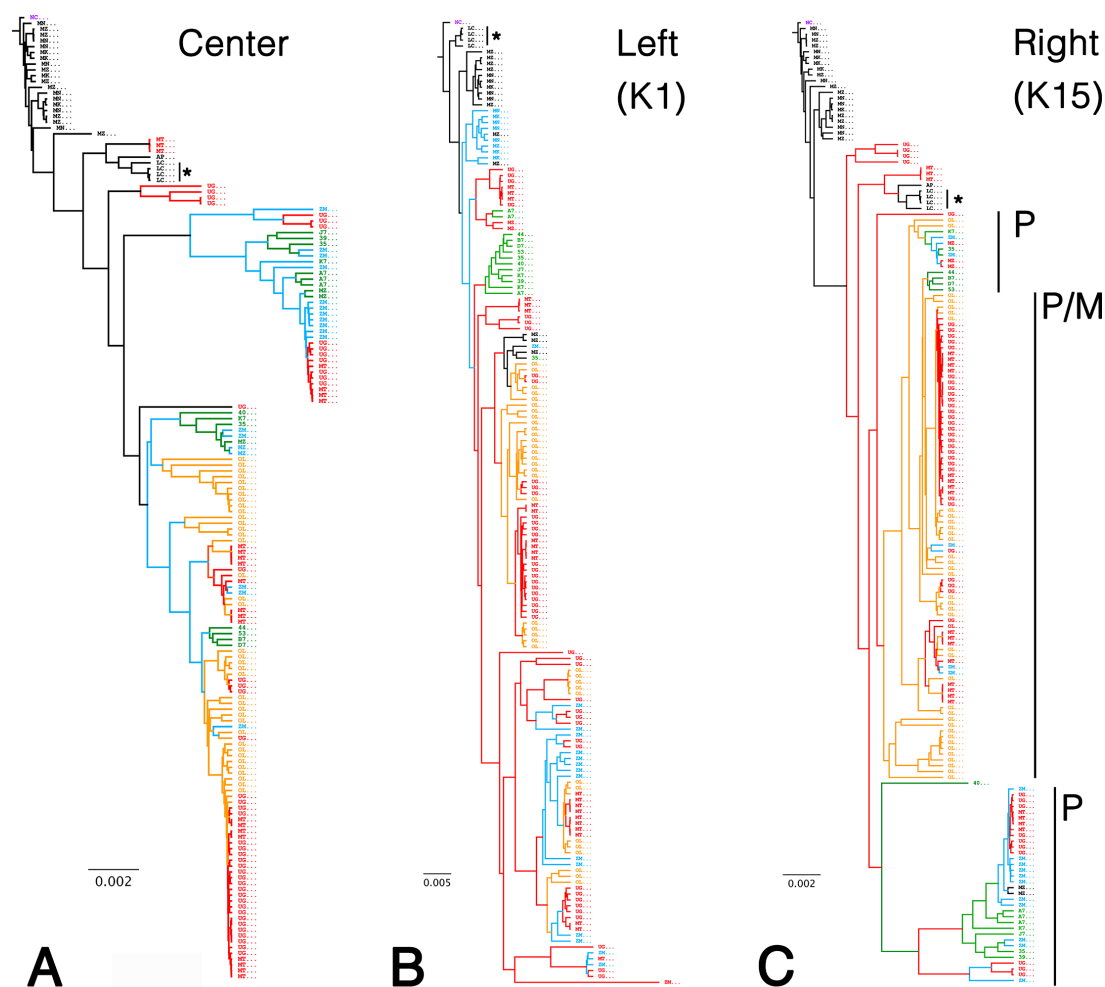
### AlphaFold predictions of proteins with shared African SNVs

Next, we modeled the structure of proteins that carried non-synonymous, shared SNVs in our phylogenetic analysis using AlphaFold2. We focused on ORFs with SNVs that resulted in an amino acid change with different chemical characteristics as these more likely result in conformational changes.

#### ORF49

The AlphaFold2 model of ORF49 was similar to the crystal structure, color coded in gray (RMSD = 0.273 Å) (Fig. 5A). The ORF49 structure consists of 12 alpha-helices bundled into two pseudo domains. In Fig. 5 and 6; Fig. S1A always shows the model with the pLDDT score is color coded, such that cooler colors (blue) indicate higher prediction confidence. The pLDDT confidence scores of the reference sequence indicate a very high confidence model as indicated by the blue color. Variability is noted in three places: first, at the N-terminal loop (red color in Fig. 5A) where the crystal structure is missing nine residues; second in the loop region on top of the structure (cyan color in Fig. 5A), which is disordered in the crystal structure; third for a stretched helix in the back (cyan color in Fig. 5A). The predicted AlphaFold model of the Malawian isolate, MZ712182 (subtype A), was identical (RMSD = 0.038 Å) (Fig. 5B) as ORF49 is highly conserved.

The sequence coverage plot indicates that AlphaFold2 used 40 sequences with a high sequence similarity across the central region of the protein, amino acids 50 to 250 and less similarity at the N and C terminus (Fig. 5C). The PAE plot (Fig. 5D) indicated a low PAE at residue pairs  $x$  and  $y$ , translating to a well-defined model of two domains (in blue) separated by a flexible region around amino acid 150 in the middle (33). The N-terminus

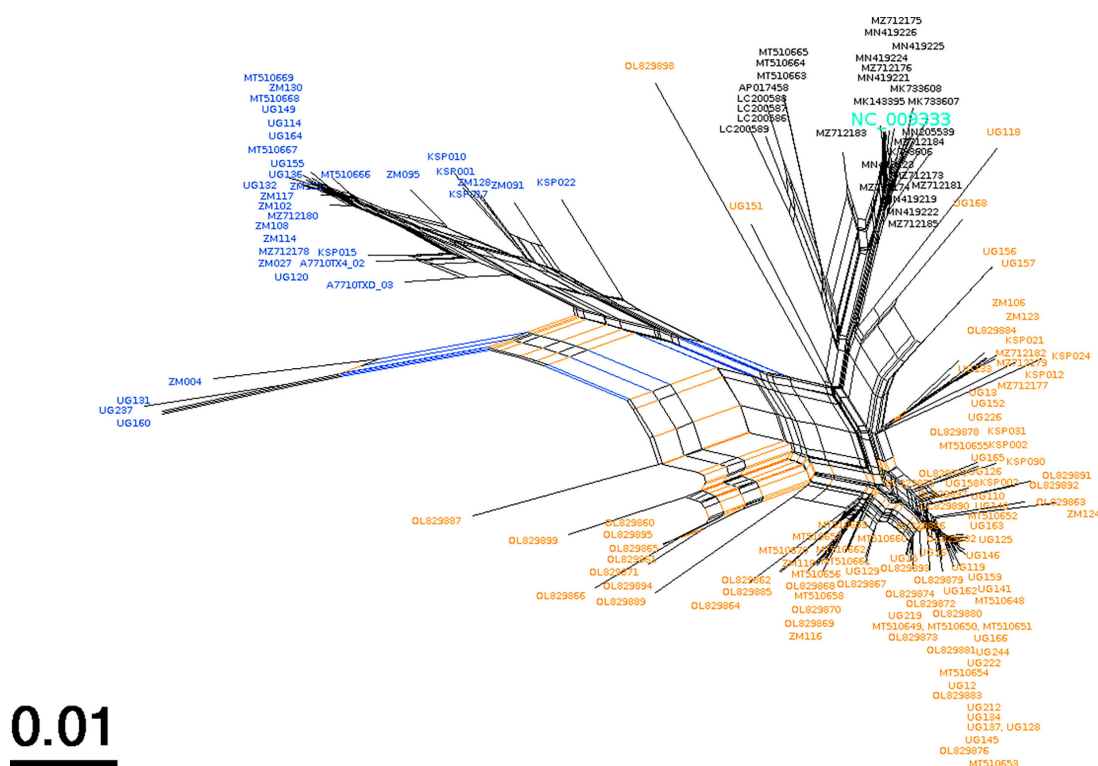


**FIG 3** Phylogenetic analysis of LUR, K1, and K15. The results of a Bayesian phylogenetic analysis of the African KSHV genomes using BEAST are shown. (A) A tree based on the LUR extracted from 164 genomes, the ML tree identifies two subtypes of KSHV genomes isolated from Africa. The country of origin is indicated by color: Malawi in green, Uganda in red, Zambia in blue, Cameroon in yellow, and USA/Europe/Japan in black. The four Japanese isolates are indicated by an asterisk. (B) A tree based on the K1 region (right side). The additional Malawian genomes described in this study are classified as the A5 subtype. (C) A tree based on the K15 gene (left side). The additional Malawian genomes described in this study are classified as the P-subtype.

also has low (red) PAE scores. The ORF49 protein reconstruction thus serves as the positive control and benchmark for subsequent predictions, as it has a ground truth in the crystal structure of the reference sequence.

### ORF22/gH

The AlphaFold2 model of gH aligned to the crystal structure (extracted from the PDB file and removing gL and the LBD) with an RMSD = 1.875 Å (544 atoms) (Fig. 6A). The model for gB was 10-fold less accurate than the model for ORF49 if measured against a single crystal structure. Some of the secondary elements are not superimposed. Several reasons may have contributed to this. Firstly, the gH crystal structure lacks 35 and 32 residues at the N-terminus and C-terminus, respectively. These residues comprise an extended loop region in the AlphaFold2 model, with predicted alpha helices and beta sheets, i.e., not unordered. Secondly, within the regions of high confidence, the misalignment of the alpha helices in the D-II domain could contribute to the high RMSD, as well. AlphaFold2 predicted an additional small alpha helix in the D-II and two small alpha helices in the D-III domain of a putative gH monomer absent in the crystal structure of gH: gL: EphA2 trimer. The model used 140 sequences to generate the MSA (Fig. 6B), with the highest



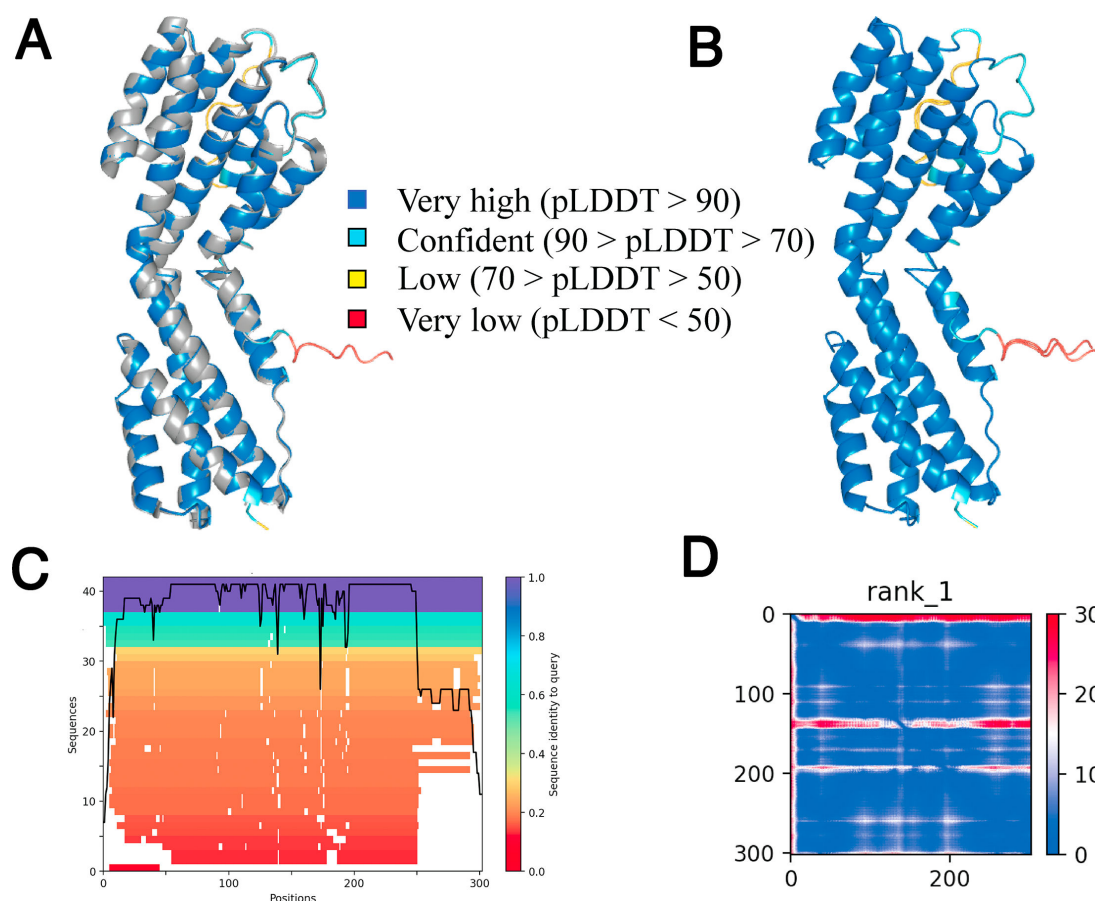
**FIG 4** SplitTree analysis. This analysis used  $n = 164$  KSHV LUR sequences to construct the ML phylogenetic tree. Darker branch shadings represent branches with higher confidence (1,000 bootstrap replicates). The Phi test for recombination found statistical evidence for recombination ( $P < 0.01$ ). A total of 1,393 informative sites were found using a window size of 100. Mean = 0.3132622742; variance = 4.70900; observed = 0.093065.

conservation in the region between amino acids 400 to the end, i.e., the C-terminal half of the protein had twice as much data as the N-terminal half. Nevertheless, the PAE indicates high confidence in the overall protein orientation (Fig. 6C).

Next, we compared the AlphaFold2 models of the reference sequence and the Malawian isolate from subtype A, MZ712182, and B. The subtype A aligned to the reference model with an RMSD = 0.137 Å (557 atoms) (Fig. 6D) and to the genome from subtype B with an RMSD = 0.201 Å (588 atoms). The model for subtype A aligned to the one of subtype B with an RMSD = 0.074 Å (570 atoms), indicating similar folding except for an additional alpha helix observed at the N-terminal and a misalignment in the loop region of D-IV. The reference sequence and Malawian isolate models aligned with the same secondary structures, except the extended helix in the D-I domain, which is shortened in subtype B and not present in subtype A. The misaligned regions include the extended loop in the D-I domain and the loop in the D-III domain (Fig. 6D).

These structural differences may be due to the four non-synonymous SNVs found in most African genomes and not in the reference strain (Fig. 6E). Two of the four SNVs are in the D-I domain. They include Val68Ala (86% frequency in subtype A; 99% frequency in subtype B) and Val104Ala (86% in subtype A; 99% in subtype B). The valine to alanine change at positions 68 and 104 are small hydrophobic amino acids, with the latter on a loop structure connecting two alpha helices. Val68Ala is on the end of an N-terminal alpha-helix. The N-terminus of ORF22 interacts with gL and the human EphA2 (39). The Val68 forms two contacts with the Leu66 residue on the alpha helix (2.2 Å and 3.0 Å). Those are likely preserved in the Ala68 variant. The Val104 forms contacts with residues on the beta-sheet; the carbonyl group of Val104 forms a hydrogen bond with the residue Ala244 (2.0 Å), and the amino group of Val104 forms a hydrogen bond with Phe243 (1.9 Å). The carbonyl and amino group of Ala104 in the Malawian genome form the same hydrogen bonds as Val104, with a slightly shorter distance of 1.8 Å between the amino





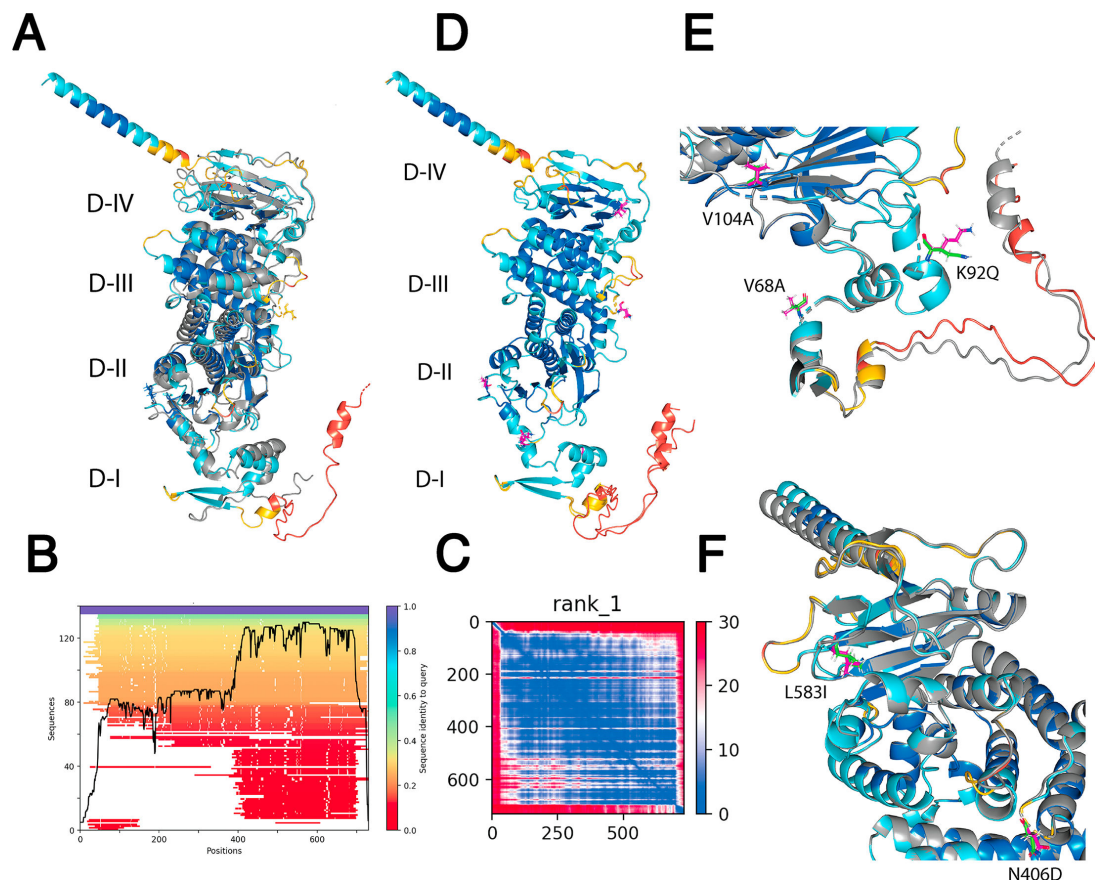
**FIG 5** Model of KSHV ORF49. (A) The ORF49 crystal structure (5IPX.pdb) in gray is aligned to the AlphaFold2 model for the reference sequence, as shown in the pLDDT color scoring. (B) The highly confident model of the ORF49 sequence from the Malawian isolates is colored in the pLDDT color scoring. (C) The sequence coverage plot indicates that over 40 sequences were used to generate the MSA and their sequence similarity to the query ORF49. The sequences were 80%–100% similar to ORF49, except at the N- and C-terminals. (D) PAE indicates that AlphaFold2 is highly confident that the predicted structure has two domains and spatial arrangement. The axes represent the residue index, i.e., position of the amino acid.

group and Phe243 compared to the reference genome, and the carbonyl group of Ala104 forms a hydrogen bond with Ala244 (2.0 Å).

The two non-synonymous SNVs found in the C-terminal include the Asn406Asp (87% in subtype A; 100% in subtype B) in the D-III domain and Leu583Ile (87% in subtype A; 100% in subtype B) in the D-IV domain (Fig. 6F). Leu583 in the reference genome forms a hydrogen bond with its amino group to Ser603 (2.8 Å) and the carbonyl group to Ala586 (2.6 Å). The Ile583 variant in the African subtypes likely maintains those interactions. These findings have immediate value, considering that these SNVs represent replication-competent, common KSHV variants. They could become hotspots for escape mutants to antibodies targeting KSHV gH or engineering targets to increase the stability of gH and thus the infectivity of KSHV bacmids.

## ORF34

There is no crystal structure of ORF34. Thus, the basis for AlphaFold2 prediction is weaker than the preceding targets (Fig. S1). The models based on the Malawian isolates differed from the reference genome and from each other. The highly conserved C-terminus of the protein aligned to subtype A, MZ712180, with an RMSD = 0.248 Å (208 atoms), and to subtype B, MZ712182, with an RMSD = 0.156 Å (220 atoms). The N-terminus differed between the African subtypes. The C-terminus of the protein aligned perfectly across all models, consistent with its conserved function and interactions. This study identified four



**FIG 6** AlphaFold2 model of KSHV ORF22/gH. (A) The ORF22 crystal structure (7czf.pdb) in gray is aligned with the reference model, as shown in the pLDDT color scoring. The domains of the structure are indicated. (B) The sequence coverage plot indicates that 140 sequences were used to generate the MSA and their sequence similarity to the query ORF22. The sequences were 40%–60% similar to ORF22, except at the N- and C-terminals, where similarity was 20%. (C) PAE indicates that AF2 is highly confident the predicted structure is multidomain and in its spatial arrangement. The axes represent the residue index, i.e., position of the amino acid. (D) The highly confident reference model of ORF22 aligned with the model of MZ712180 of subtype A and was colored in the pLDDT color scoring. The structures are almost identical except for the N- and C-terminal misalignment and the additional alpha helix at the N-terminal. (E) The two non-synonymous SNVs identified in the D-I domain are V68A and V104A. (F) The two non-synonymous SNVs are found in the D-III and D-IV domains, N406 and L583I, respectively (green line residue represents the reference sequence, and pink stick residue represents the Malawian sequence).

common, non-synonymous variants in ORF34. The Ala225Ser (82% in subtype A; 98% in subtype B) was present in almost all African isolates. Leu64Gln (81% frequency) and Glu96Lys (85% frequency) are located in the N-terminus and are only present in subtype A and Cys61Ser (94% frequency) in subtype B. These were located on loop regions of the protein and modeled with low confidence. Cys61Ser in subtype B may be responsible for the loss of the terminal alpha helix, which is present in both subtype A and the reference genome. If true, this represents a drastic change in the protein structure. The conserved amino acids in the N-terminal of ORF34 remain conserved in all genomes described here.

Exploring every non-synonymous SNV in every one of the 80 proteins of KSHV by AlphaFold2 is not feasible. Also, the sequence space of KSHV variants is still limited, as is structure prediction accuracy. Nevertheless, these examples provide a blueprint for high-throughput evolution to structure experiments.

## DISCUSSION

This study added 13 complete KSHV genomes to the growing collection of publicly available KSHV genomes isolated from KS endemic areas in recent calendar time (6, 22, 32, 45–48, 64, 65, 81, 82). It extends community efforts to assemble a comprehensive collection of complete viral genomes representing KSHV diversity today. The samples

were collected from HIV-positive adult KS patients at study enrollment, i.e., prior to any KS treatment (31). Eight samples were from plasma, and seven were from KS tumor tissue. One pair of samples was matched, i.e., obtained from different compartments for the same participant.

Furthermore, this study added KSHV<sup>BC-1</sup>, KSHV<sup>BCBL-1</sup>, KSHV<sup>BC-3</sup>, and KSHV<sup>JSC-1</sup> genomes as ascertained by long-read sequencing of latent virus and *de novo* assembly. The initial genomes were remarkably error-free already, given that they were assembled by Sanger Sequencing at ~12× coverage (22). We also obtained the complete human genomes for these cell lines, extending our earlier studies (83).

Even though the study isolated DNA from 137 plasma samples and 75 KS FFPE blocks, PCR-based targeted enrichment and sequencing succeeded only for 6% of the plasma and 8% of the FFPE blocks. Success was defined as a sequencing depth sufficient to identify SNVs with high confidence across the entire genome. Several explanations exist for the limited yield of complete genomes from “field samples.”

First, concerning sequencing from FFPE blocks. Preserving samples collected under routine clinical practice conditions in low- and middle-income countries will always be sub-optimal. Improper (wrong pH) storage or over-fixation (due to lack of personnel) of tumor biopsies is associated with DNA degradation and fragmentation. While DNA fragmentation is less a problem for diagnostic PCR and limited sequencing of individual genes, it dramatically diminishes the quality of whole genome assembly, which at its core assumes even and equal coverage across the targeted genome.

Second, concerning sequencing from plasma. The DNA copy numbers for KSHV are an order of magnitude lower than saliva or lesion-associated genome copy numbers (66, 84–86). The KSHV plasma DNA copy number is typically lower than the EBV or HCMV DNA copy numbers in affected individuals, rarely rising above five log<sub>10</sub> copies/mL. Extraordinarily high plasma genome copy numbers usually indicate the presence of concurrent KAD, such as MCD or KICS (64, 65). This difference in DNA copy number may explain why GenBank has more complete EBV (*n* ~1,375) than KSHV (~255) genomes. A higher level of performance is required to attain complete KSHV genomes from KS patient material than from EBV or HCMV carriers.

The KSHV genomes used for phylogenetic analyses were from GenBank. Acceptance in GenBank assures a certain level of quality control but only minimal association with metadata. The entries or derivatives are publicly available and indefinitely preserved. The GenBank identifiers are unique, and the coding regions are annotated to match the KSHV reference sequence. Additional sequence information for KSHV exists as unassembled reads in SRA archives, e.g., bio project PRJEB2768. Some of this information was used to assemble some KSHV genomes for phylogenetic analysis; however, since we did not do the sequencing, it remains for the authors to submit those sequences to GenBank.

The phylogenetic analysis presented here confirmed our prior discovery of two distinct KSHV lineages co-circulating in KSHV endemic countries today (32). This updated analysis gained greater power and granularity by adding data generously shared by others (6, 22, 32, 45–48, 64, 65, 81). The extended phylogenetic analysis uncovered several non-synonymous SNVs common to each clade or prevalent within particular countries. The extended phylogenetic analyses separate current African KSHV isolates from earlier European and US isolates.

The four Japanese KSHV strains clustered separately. These isolates are from classic KS cases and are classified as the C subtype of K1 and M based on their K15 allele. The virus sequences have approximately 500 non-synonymous mutations compared to the reference strain. Most of those are shared among them; however, there is a distinction between the three genomes from the Miyako islands and the one from mainland Japan. In an easily overlooked publication, Awazawa et al. reported that these genomes were collected from classic KS patients living on the Miyako Islands in Okinawa Prefecture at the Southernmost tip of Japan (74). On the Miyako Islands, KSHV seroprevalence is 15.4%, which is significantly above mainland Japan, the US, or Northern Europe. The

separation in sequence space for the Japanese KSHV strains supports our contention of continuous KSHV sequence drift, here uncovered by separation in geographical space.

Geographic isolation has been one of the fundamental mechanisms of speciation ever since Darwin documented different finch species on different islands within the Galápagos Archipelago (reviewed in reference [87]). It applies to KSHV evolution as well. The herpesviruses evolved with human migration as early as the Bronze Age (88, 89). Then, geography separated the KSHV strains circulating in areas of moderate to high prevalence (15%–75%) in Japan, Africa, Europe (mostly the Levant), and the Uyghur populations in Xinjiang, China (90). A limitation of this conjecture is that we only have enough sequence isolates from SSA to exclude sample bias and founder effects.

For the high-coverage genomes, low-frequency SNVs could be identified with high confidence. This, together with the formal detection of recombination events, supports the notion of reinfection, intra-host evolution, or concurrent reactivation events within a lesion, a person, or a saliva sample (46). How frequent these are and whether they are linked to particular environmental, clinical conditions, or age is subject to further study. None of the recombinant KSHV genome sequences in a mixed infection lesion have been evaluated with regard to replication competency. Hence, there exists the possibility that these represent defective rearrangements.

The severe paucity of high-resolution structures for most KSHV proteins remains a barrier to inferring biological consequences of sequence variation. Machine learning programs have recently pushed the boundaries of protein structure prediction. These programs provided models to put some viral SNVs into a structural context. AlphaFold2 was used here due to its superiority over other algorithms in determining protein structures (33). There are limitations specific to each protein, however. Each prediction has to be considered based on the underlying evidence to build the model. Some KSHV ORFs are homologous to other herpesvirus ORFs, suggesting a conserved fold. For some viral ORFs (KSHV and other herpesviruses), CryoEM or X-ray structures were incorporated by AlphaFold2. For other viral ORFs, there is little to no information beyond general folding tendencies to derive a structure. We consider such modeling approaches as hypothesis-generating, which has value in itself. Future experimental verification of KSHV protein structures is urgently needed. It is unclear how adding additional, more diverse KSHV protein sequences will affect structure prediction. Will AlphaFold get worse, begin to “hallucinate” even, and how would we find out? We opine that obtaining more high-resolution structures of KSHV protein variants, i.e., ground truth data, is as important as obtaining more genomic sequences.

Several SNVs mapped to KSHV surface proteins and regulatory proteins. Mapping SNVs to the model for gH may suggest stabilizing mutations for crystallography, and it may inform vaccine design. By severely straining credulity, one could argue that some SSA-specific and conserved SNVs represent receptor usage patterns based on ephrin alleles prevalent in endemic areas but not in US and European populations (91, 92). This would explain the non-universal distribution of KSHV, which is unusual among human herpesviruses.

## ACKNOWLEDGMENTS

First and foremost, we thank the participants without whom our study would not have been possible.

This work was supported by public health service grants CA254564, CA019014, and DE018304 to D.P.D. and CA16086 to the UNC Lineberger Comprehensive Cancer Center. R.M. is a fellow of the AIDS Malignancy Consortium (CA121947).

## AUTHOR AFFILIATION

<sup>1</sup>Lineberger Comprehensive Cancer Center and Department of Microbiology and Immunology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

## AUTHOR ORCIDs

Dirk P. Dittmer  <http://orcid.org/0000-0003-4968-5656>

## FUNDING

Funder	Grant(s)	Author(s)
National Institute of Dental and Craniofacial Research	DE018304	Dirk P. Dittmer
National Cancer Institute	CA16086	Dirk P. Dittmer
National Cancer Institute	CA121947	Razia Moorad

## AUTHOR CONTRIBUTIONS

Razia Moorad, Data curation, Formal analysis, Investigation, Writing – original draft | Alice Peng, Data curation, Formal analysis, Software | Justin Landis, Data curation, Software | Linda J. Pluta, Investigation, Resources | Patricio Cano, Investigation, Resources | Angelica Juarez, Data curation, Methodology, Resources, Validation | Dirk P. Dittmer, Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing – review and editing

## ETHICS APPROVAL

The study was approved by the Malawi National Health Science and Research Committee (NHSRC) and the Institutional Review Board (IRB) at the University of North Carolina at Chapel Hill.

## ADDITIONAL FILES

The following material is available [online](#).

## Supplemental Material

**Figure S1 (JVI01950-24-S0001.docx).** AlphaFold2 model of KSHV ORF34.

## REFERENCES

1. Openshaw MR, Gervasi E, Fulgenzi CAM, Pinato DJ, Dalla Pria A, Bower M. 2023. Taxonomic reclassification of Kaposi sarcoma identifies disease entities with distinct immunopathogenesis. *J Transl Med* 21:283. <https://doi.org/10.1186/s12967-023-04130-6>
2. Lage SL, Ramaswami R, Rocco JM, Rupert A, Davis DA, Lurain K, Manion M, Whitby D, Yarchoan R, Sereti I. 2024. Inflammasome activation in patients with Kaposi sarcoma herpesvirus-associated diseases. *Blood* 144:1496–1507. <https://doi.org/10.1182/blood.2024024144>
3. Hansen ME, Mangusan R, Lurain K, Odeny T, George J, Lu C, Manion M, Widell A, Ekwede I, Whitby D, Gulley JL, Kadri SS, Elinoff JM, Barochia A, Torabi-Parizi P, Uldrick TS, Yarchoan R, Ramaswami R. 2022. Characteristics of patients admitted to the ICU with Kaposi sarcoma herpesvirus-associated diseases. *AIDS* 36:1969–1978. <https://doi.org/10.1097/QAD.0000000000003333>
4. Cesarman E, Damania B, Krown SE, Martin J, Bower M, Whitby D. 2019. Kaposi sarcoma. *Nat Rev Dis Primers* 5:9. <https://doi.org/10.1038/s41572-019-0060-9>
5. Damania B, Dittmer DP. 2023. Today's Kaposi sarcoma is not the same as it was 40 years ago, or is it? *J Med Virol* 95:e28773. <https://doi.org/10.1002/jmv.28773>
6. Caro-Vegas C, Peng A, Juarez A, Silverstein A, Kamiyango W, Villiera J, McAtee CL, Mzikamanda R, Tomoka T, Peckham-Gregory EC, Moorad R, Kovarik CL, Campbell LR, Mehta PS, Kazembe PN, Allen CE, Scheurer ME, Ozuah NW, Dittmer DP, El-Mallawany NK. 2023. Pediatric HIV+ Kaposi sarcoma exhibits clinical, virological, and molecular features different from the adult disease. *JCI Insight* 8:e167854. <https://doi.org/10.1172/jci.insight.167854>
7. Brayfield BP, Kankasa C, West JT, Muyanga J, Bhat G, Klaskala W, Mitchell CD, Wood C. 2004. Distribution of Kaposi sarcoma-associated herpesvirus/human herpesvirus 8 in maternal saliva and breast milk in Zambia: implications for transmission. *J Infect Dis* 189:2260–2270. <https://doi.org/10.1086/421119>
8. Poppe LK, Kankasa C, Wood C, West JT. 2020. Relationships between maternal antibody responses and early childhood infection with Kaposi sarcoma-associated herpesvirus. *J Infect Dis* 222:1723–1730. <https://doi.org/10.1093/infdis/jiaa288>
9. Nalwoga A, Cose S, Wakeham K, Miley W, Ndibazza J, Drakeley C, Elliott A, Whitby D, Newton R. 2015. Association between malaria exposure and Kaposi's sarcoma-associated herpes virus seropositivity in Uganda. *Trop Med Int Health* 20:665–672. <https://doi.org/10.1111/tmi.12464>
10. Sabourin KR, Daud I, Ogolla S, Labo N, Miley W, Lamb M, Newton R, Whitby D, Rochford R. 2021. Malaria is associated with Kaposi sarcoma-associated herpesvirus seroconversion in a cohort of Western Kenyan children. *J Infect Dis* 224:303–311. <https://doi.org/10.1093/infdis/jiaa740>
11. Oluoch PO, Oduor CI, Forconi CS, Ong'echa JM, Münz C, Dittmer DP, Bailey JA, Moormann AM. 2020. Kaposi Sarcoma-associated herpesvirus infection and endemic burkitt lymphoma. *J Infect Dis* 222:111–120. <https://doi.org/10.1093/infdis/jiaa060>
12. Motlhale M, Sitas F, Bradshaw D, Chen WC, Singini MG, de Villiers CB, Lewis CM, Muchengeti M, Waterboer T, Mathew CG, Newton R, Singh E. 2022. Epidemiology of Kaposi's sarcoma in sub-saharan Africa. *Cancer Epidemiol* 78:102167. <https://doi.org/10.1016/j.canep.2022.102167>
13. Peprah S, Engels EA, Horner MJ, Monterosso A, Hall HI, Johnson AS, Pfeiffer RM, Shiels MS. 2021. Kaposi sarcoma incidence, burden, and prevalence in United States people with HIV, 2000-2015. *Cancer Epidemiol Biomarkers Prev* 30:1627–1633. <https://doi.org/10.1158/1055-9965.EPI-21-0008>



14. Salyards M, Nijhawan AE, Kuo J, Knights SM, Lazarte S, Labo N, Miley W, Whitby D, Hwang LY, Kornberg AW, Fujimoto K, Chiao EY. 2024. Prevalence, incidence, and predictors of Kaposi sarcoma-associated herpesvirus infection among young men who have sex with men in the southern United States. *J Infect Dis* 229:1387–1392. <https://doi.org/10.1093/infdis/jiad384>
15. Hosseiniour MC, Kang M, Krown SE, Bukuru A, Umbleja T, Martin JN, Orem J, Godfrey C, Hoagland B, Mwelase N, Langat D, Nyirenda M, MacRae J, Borok M, Samaneka W, Moses A, Mngqibisa R, Busakhala N, Martínez-Maza O, Ambinder R, Dittmer DP, Nokta M, Campbell TB, Team AA-R-K. 2018. As-needed vs immediate etoposide chemotherapy in combination with antiretroviral therapy for mild-to-moderate AIDS-associated Kaposi sarcoma in resource-limited settings: A5264/AMC-067 Randomized Clinical Trial. *Clin Infect Dis* 67:251–260. <https://doi.org/10.1093/cid/ciy044>
16. Krown SE, Lee JY, Dittmer DP, ConsortiumAM. 2008. More on HIV-associated Kaposi's sarcoma. *N Engl J Med* 358:535–536. <https://doi.org/10.1056/NEJMc072994>
17. 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. <https://doi.org/10.1056/NEJMc070508>
18. Burger H, Ismail Z, Taljaard JJ. 2018. Establishing a multidisciplinary AIDS-associated Kaposi's sarcoma clinic: patient characteristics, management and occurrence of Kaposi's sarcoma in patients living with human immunodeficiency virus (HIV) and on antiretroviral therapy, despite suppressed HIV viremia. *Clin Infect Dis* 70:2435–2438. <https://doi.org/10.1093/cid/ciz762>
20. Knights SM, Lazarte SM, Kainthla R, Chiao EY, Nijhawan AE. 2022. Mortality disparities among patients with HIV-associated Kaposi's sarcoma in the southern United States. *AIDS* 36:721–728. <https://doi.org/10.1097/QAD.0000000000003155>
21. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. 2021. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 71:209–249. <https://doi.org/10.3322/caac.21660>
22. Russo JJ, Bohenzky RA, Chien MC, Chen J, Yan M, Maddalena D, Parry JP, Peruzzi D, Edelman IS, Chang Y, Moore PS. 1996. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). *Proc Natl Acad Sci USA* 93:14862–14867. <https://doi.org/10.1073/pnas.93.25.14862>
23. Cesarman E, Moore PS, Rao PH, Inghirami G, Knowles DM, 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. <https://doi.org/10.1182/blood.V86.7.2708.2708>
24. Cannon JS, Ciuffo D, Hawkins AL, Griffin CA, Borowitz MJ, Hayward GS, Ambinder RF. 2000. A new primary effusion lymphoma-derived cell line yields a highly infectious Kaposi's sarcoma herpesvirus-containing supernatant. *J Virol* 74:10187–10193. <https://doi.org/10.1128/jvi.74.21.10187-10193.2000>
25. Renne R, Zhong W, Herndier B, McGrath M, Abbey N, Kedes D, Ganem D. 1996. Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. *Nat Med* 2:342–346. <https://doi.org/10.1038/nm0396-342>
26. Brulois KF, Chang H, Lee A-Y, Ensser A, Wong L-Y, Toth Z, Lee SH, Lee H-R, Myoung J, Ganem D, Oh T-K, Kim JF, Gao S-J, Jung JU. 2012. Construction and manipulation of a new Kaposi's sarcoma-associated herpesvirus bacterial artificial chromosome clone. *J Virol* 86:9708–9720. <https://doi.org/10.1128/JVI.01019-12>
27. Zhou FC, Zhang YJ, Deng JH, Wang XP, Pan HY, Hettler E, Gao SJ. 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. <https://doi.org/10.1128/jvi.76.12.6185-6196.2002>
28. Gatherer D, Depledge DP, Hartley CA, Szpara ML, Vaz PK, Benkő M, Brandt CR, Bryant NA, Dastjerdi A, Doszpoly A, Gompels UA, Inoue N, Jarosinski KW, Kaul R, Lacoste V, Norberg P, Origgi FC, Orton RJ, Pellett PE, Schmid DS, Spatz SJ, Stewart JP, Trimpert J, Waltzek TB, Davison AJ. 2021. ICTV virus taxonomy profile: herpesviridae 2021. *J Gen Virol* 102. <https://doi.org/10.1099/jgv.0.001673>
29. Hayward GS. 1999. KSHV strains: the origins and global spread of the virus. *Semin Cancer Biol* 9:187–199. <https://doi.org/10.1006/scbi.1998.0116>
30. Poole LJ, Zong JC, Ciuffo DM, Alcendor DJ, Cannon JS, Ambinder R, Orenstein JM, Reitz MS, Hayward GS. 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. <https://doi.org/10.1128/JVI.73.8.6646-6660.1999>
31. Moorad R, Kasonkanji E, Gumulira J, Gondwe Y, Dewey M, Pan Y, Peng A, Pluta LJ, Kudowa E, Nyasosela R, Tomoka T, Tweya H, Heller T, Gugsa S, Phiri S, Moore DT, Damania B, Painschab M, Hosseiniour MC, Dittmer DP. 2023. A prospective cohort study identifies two types of HIV+ Kaposi sarcoma lesions: proliferative and inflammatory. *Int J Cancer* 153:2082–2092. <https://doi.org/10.1002/ijc.34689>
32. Moorad R, Juarez A, Landis JT, Pluta LJ, Perkins M, Cheves A, Dittmer DP. 2022. Whole-genome sequencing of Kaposi sarcoma-associated herpesvirus (KSHV/HHV8) reveals evidence for two African lineages. *Virology* 568:101–114. <https://doi.org/10.1016/j.virol.2022.01.005>
33. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature New Biol* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
34. González CM, Wong EL, Bowser BS, Hong GK, Kenney S, Damania B. 2006. Identification and characterization of the ORF49 protein of Kaposi's sarcoma-associated herpesvirus. *J Virol* 80:3062–3070. <https://doi.org/10.1128/JVI.80.6.3062-3070.2006>
35. Lee S, Cho HJ, Park JJ, Kim YS, Hwang S, Sun R, Song MJ. 2007. The ORF49 protein of murine gammaherpesvirus 68 cooperates with RTA in regulating virus replication. *J Virol* 81:9870–9877. <https://doi.org/10.1128/JVI.00001-07>
36. Hew K, Veerappan S, Sim D, Cornvik T, Nordlund P, Dahlroth SL. 2017. Structure of the open reading frame 49 protein encoded by Kaposi's sarcoma-associated herpesvirus. *J Virol* 91:e01947-16. <https://doi.org/10.1128/JVI.01947-16>
37. Muniraju M, Mutsunguma LZ, Foley J, Escalante GM, Rodríguez E, Nabiee R, Totonchy J, Mulama DH, Nyagol J, Wussow F, Barasa AK, Brehm M, Ogembo JG. 2019. Kaposi sarcoma-associated herpesvirus glycoprotein H is indispensable for infection of epithelial, endothelial, and fibroblast cell types. *J Virol* 93:e00630-19. <https://doi.org/10.1128/JVI.00630-19>
38. Zhu FX, Chong JM, Wu L, Yuan Y. 2005. Virion proteins of Kaposi's sarcoma-associated herpesvirus. *J Virol* 79:800–811. <https://doi.org/10.1128/JVI.79.2.800-811.2005>
39. Su C, Wu L, Chai Y, Qi J, Tan S, Gao GF, Song H, Yan J. 2020. Molecular basis of EphA2 recognition by gHgL from gammaherpesviruses. *Nat Commun* 11:5964. <https://doi.org/10.1038/s41467-020-19617-9>
40. Kobayashi R, Abe M, Oguri K, Torikai M, Nishimura T, Mori H, Koshizuka T, Inoue N. 2018. Analysis of relationships between polymorphisms in the genes encoding the pentameric complex and neutralization of clinical cytomegalovirus isolates. *Vaccine* 36:5983–5989. <https://doi.org/10.1016/j.vaccine.2018.08.054>
41. Zehner M, Alt M, Ashurov A, Goldsmith JA, Spies R, Weiler N, Lerma J, Gieselmann L, Stöhr D, Gruell H, et al. 2023. Single-cell analysis of memory B cells from top neutralizers reveals multiple sites of vulnerability within HCMV Trimer and pentamer. *Immunity* 56:2602–2620. <https://doi.org/10.1016/j.immuni.2023.10.009>
42. Nishimura M, Watanabe T, Yagi S, Yamanaka T, Fujimuro M. 2017. Kaposi's sarcoma-associated herpesvirus ORF34 is essential for late gene expression and virus production. *Sci Rep* 7:329. <https://doi.org/10.1038/s41598-017-00401-7>
43. Brulois K, Wong LY, Lee HR, Sivasdas P, Ensser A, Feng P, Gao SJ, Toth Z, Jung JU. 2015. Association of Kaposi's sarcoma-associated herpesvirus ORF31 with ORF34 and ORF24 is critical for late gene expression. *J Virol* 89:6148–6154. <https://doi.org/10.1128/JVI.00272-15>
44. Davis ZH, Hesser CR, Park J, Glaunsinger BA. 2016. Interaction between ORF24 and ORF34 in the Kaposi's sarcoma-associated herpesvirus late gene transcription factor complex is essential for viral late gene expression. *J Virol* 90:599–604. <https://doi.org/10.1128/JVI.02157-15>
45. Olp LN, Jeanniard A, Marimo C, West JT, 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. <https://doi.org/10.1128/JVI.01712-15>



46. Santiago JC, Goldman JD, Zhao H, Pankow AP, Okuku F, Schmitt MW, Chen LH, Hill CA, Casper C, Phipps WT, Mullins JI. 2021. Intra-host changes in Kaposi sarcoma-associated herpesvirus genomes in Ugandan adults with Kaposi sarcoma. *PLoS Pathog* 17:e1008594. <https://doi.org/10.1371/journal.ppat.1008594>
47. Sallah N, Palser AL, Watson SJ, 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. <https://doi.org/10.1093/infdis/jiy427>
48. Marshall VA, Fisher NC, Goodman CA, Cornejo Castro EM, Liu I, Khanal S, Holdridge BM, Thorpe AL, Labo N, Stolka KB, Hemingway-Foday JJ, Abassora M, N'Dom P, Smith JS, Sallah N, Palser AL, Kellam P, Keele BF, Whitby D. 2022. Systematic analysis of Kaposi's sarcoma (KS)-associated herpesvirus genomes from a KS case-control study in Cameroon: evidence of dual infections but no association between viral sequence variation and KS risk. *Int J Cancer* 151:1127–1141. <https://doi.org/10.1002/ijc.34136>
49. Glenn M, Rainbow L, Auradé F, Davison A, Schulz TF. 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. <https://doi.org/10.1128/JVI.73.8.6953-6963.1999>
50. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
51. Drummond AJ, Rambaut A. 2007. BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214. <https://doi.org/10.1186/1471-2148-7-214>
52. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST. *Mol Biol Evol*. <https://doi.org/10.1093/molbev/mss075>
53. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
54. Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23:254–267. <https://doi.org/10.1093/molbev/msj030>
55. 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. <https://doi.org/10.1128/JVI.05071-11>
56. Vieira J, O'Hearn PM. 2004. Use of the red fluorescent protein as a marker of Kaposi's sarcoma-associated herpesvirus lytic gene expression. *Virology* 325:225–240. <https://doi.org/10.1016/j.virol.2004.03.049>
57. Abere B, Mamo TM, Hartmann S, Samarina N, Hage E, Rückert J, Hotop S-K, Büsche G, Schulz TF. 2017. The Kaposi's sarcoma-associated herpesvirus (KSHV) non-structural membrane protein K15 is required for viral lytic replication and may represent a therapeutic target. *PLoS Pathog* 13:e1006639. <https://doi.org/10.1371/journal.ppat.1006639>
58. Knickmann J, Staliunaite L, Puhach O, Ostermann E, Günther T, Nichols J, Jarvis MA, Voigt S, Grundhoff A, Davison AJ, Brune W. 2024. A simple method for rapid cloning of complete herpesvirus genomes. *Cell Rep Methods* 4:100696. <https://doi.org/10.1016/j.crmeth.2024.100696>
59. Yakushko Y, Hackmann C, Günther T, Rückert J, Henke M, Koste L, Alkharsah K, Bohne J, Grundhoff A, Schulz TF, 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. <https://doi.org/10.1128/JVI.02068-10>
60. Bigi R, Landis JT, An H, Caro-Vegas C, Raab-Traub N, Dittmer DP. 2018. Epstein-Barr virus enhances genome maintenance of Kaposi sarcoma-associated herpesvirus. *Proc Natl Acad Sci USA* 115:E11379–E11387. <https://doi.org/10.1073/pnas.1810128115>
61. Renne R, Lagunoff M, Zhong W, Ganem D. 1996. The size and conformation of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in infected cells and virions. *J Virol* 70:8151–8154. <https://doi.org/10.1128/JVI.70.11.8151-8154.1996>
62. Ballestas ME, Chatis PA, Kaye KM. 1999. Efficient persistence of extrachromosomal KSHV DNA mediated by latency-associated nuclear antigen. *Science* 284:641–644. <https://doi.org/10.1126/science.284.5414.641>
63. Chiu YF, Sugden AU, Fox K, Hayes M, Sugden B. 2017. Kaposi's sarcoma-associated herpesvirus stably clusters its genomes across generations to maintain itself extrachromosomally. *J Cell Biol* 216:2745–2758. <https://doi.org/10.1083/jcb.201702013>
64. Tamburro KM, Yang D, Poisson J, Fedorow Y, Roy D, Lucas A, Sin S-H, Malouf N, Moylan V, Damania B, Moll S, van der Horst C, Dittmer DP. 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. <https://doi.org/10.1016/j.virol.2012.08.014>
65. Caro-Vegas C, Sellers S, Host KM, Seltzer J, Landis J, Fischer WA II, Damania B, Dittmer DP. 2020. Runaway Kaposi sarcoma-associated herpesvirus replication correlates with systemic IL-10 levels. *Virology* 539:18–25. <https://doi.org/10.1016/j.virol.2019.10.002>
66. Nalwoga A, Nakibuule M, Marshall V, Miley W, Labo N, Cose S, Whitby D, Newton R. 2020. Risk factors for Kaposi's sarcoma-associated herpesvirus DNA in blood and in saliva in rural Uganda. *Clin Infect Dis* 71:1055–1062. <https://doi.org/10.1093/cid/ciz916>
67. Phipps W, Saracino M, Selke S, Huang ML, Jaoko W, Mandaliya K, Wald A, Casper C, McClelland RS. 2014. Oral HHV-8 replication among women in Mombasa, Kenya. *J Med Virol* 86:1759–1765. <https://doi.org/10.1002/jmv.23941>
68. Casper C, Krantz E, Selke S, Kuntz SR, Wang J, Huang M-L, Pauk JS, Corey L, Wald A. 2007. Frequent and asymptomatic oropharyngeal shedding of human herpesvirus 8 among immunocompetent men. *J INFECT DIS* 195:30–36. <https://doi.org/10.1086/509621>
69. Lagunoff M, Ganem D. 1997. The structure and coding organization of the genomic termini of Kaposi's sarcoma-associated herpesvirus. *Virology* 236:147–154. <https://doi.org/10.1006/viro.1997.8713>
70. Wakeham K, Webb EL, Sebina I, Nalwoga A, Muhangi L, Miley W, Johnston WT, Ndibazza J, Whitby D, Newton R, Elliott AM. 2013. Risk factors for seropositivity to Kaposi sarcoma-associated herpesvirus among children in Uganda. *J Acquir Immune Defic Syndr* 63:228–233. <https://doi.org/10.1097/QAI.0b013e31828a7056>
71. Mbulaiteye SM, Pfeiffer RM, Engels EA, Marshall V, Bakaki PM, Owor AM, Ndugwa CM, Katongole-Mbidde E, Goedert JJ, Biggar RJ, Whitby D. 2004. Detection of Kaposi sarcoma-associated herpesvirus DNA in saliva and buffy-coat samples from children with sickle cell disease in Uganda. *J Infect Dis* 190:1382–1386. <https://doi.org/10.1086/424489>
72. Mbulaiteye S, Marshall V, Bagni RK, Wang CD, Mbisa G, Bakaki PM, Owor AM, Ndugwa CM, Engels EA, Katongole-Mbidde E, Biggar RJ, Whitby D. 2006. Molecular evidence for mother-to-child transmission of Kaposi sarcoma-associated herpesvirus in Uganda and K1 gene evolution within the host. *J Infect Dis* 193:1250–1257. <https://doi.org/10.1086/503052>
73. Zong JC, Metroka C, Reitz MS, Nicholas J, Hayward GS. 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. <https://doi.org/10.1128/JVI.71.3.2505-2511.1997>
74. Awazawa R, Utsumi D, Katano H, 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* 216:850–858. <https://doi.org/10.1093/infdis/jix424>
75. Cook PM, Whitby D, Calabro ML, Luppi M, Kakoola DN, Hjalgrim H, Ariyoshi K, Ensolli B, Davison AJ, Schulz TF. 1999. Variability and evolution of Kaposi's sarcoma-associated herpesvirus in Europe and Africa. international collaborative group. *AIDS* 13:1165–1176. <https://doi.org/10.1097/00002030-199907090-00004>
76. Lopes A de O, Spitz N, Reis C de S, de Paula VS. 2021. Update of the global distribution of human gammaherpesvirus 8 genotypes. *Sci Rep* 11:7640. <https://doi.org/10.1038/s41598-021-87038-9>
77. Isaacs T, Abera AB, Muloiwa R, Katz AA, Todd G. 2016. Genetic diversity of HHV8 subtypes in South Africa: a5 subtype is associated with extensive disease in AIDS-KS. *J Med Virol* 88:292–303. <https://doi.org/10.1002/jmv.24328>
78. Lacoste V, Judde JG, Brière J, Tulliez M, Garin B, Kassa-Kelembho E, Morvan J, Coupié P, Clyti E, Forteza Vila J, Rio B, Delmer A, Maucière P, Gessain A. 2000. Molecular epidemiology of human herpesvirus 8 in Africa: both B and A5 K1 genotypes, as well as the M and P genotypes of K14.1/K15 loci, are frequent and widespread. *Virology* 278:60–74. <https://doi.org/10.1006/viro.2000.0629>

79. Hayward GS, Zong JC. 2007. Modern evolutionary history of the human KSHV genome. *Curr Top Microbiol Immunol* 312:1–42. [https://doi.org/10.1007/978-3-540-34344-8\\_1](https://doi.org/10.1007/978-3-540-34344-8_1)
80. Kloepper TH, Huson DH. 2008. Drawing explicit phylogenetic networks and their integration into splitstree. *BMC Evol Biol* 8:22. <https://doi.org/10.1186/1471-2148-8-22>
81. Cornejo Castro EM, Marshall V, Lack J, Lurain K, Immonen T, Labo N, Fisher NC, Ramaswami R, Polizzotto MN, Keele BF, Yarchoan R, Uldrick TS, Whitby D. 2020. Dual infection and recombination of Kaposi sarcoma herpesvirus revealed by whole-genome sequence analysis of effusion samples. *Virus Evol* 6:veaa047. <https://doi.org/10.1093/ve/veaa047>
82. Marshall VA, Cornejo Castro EM, Goodman CA, Labo N, Liu I, Fisher NC, Moore KN, Nair A, Immonen T, Keele BF, Polizzotto MN, Uldrick TS, Mu Y, Saswat T, Krug LT, McBride KM, Lurain K, Ramaswami R, Yarchoan R, Whitby D. 2024. Sequencing of Kaposi's sarcoma herpesvirus (KSHV) genomes from persons of diverse ethnicities and provenances with KSHV-associated diseases demonstrate multiple infections, novel polymorphisms, and low intra-host variance. *PLoS Pathog* 20:e1012338. <https://doi.org/10.1371/journal.ppat.1012338>
83. Yang D, Chen W, Xiong J, Sherrod CJ, Henry DH, Dittmer DP. 2014. Interleukin 1 receptor-associated kinase 1 (IRAK1) mutation is a common, essential driver for Kaposi sarcoma herpesvirus lymphoma. *Proc Natl Acad Sci USA* 111:E4762–8. <https://doi.org/10.1073/pnas.1405423111>
84. Hosseinipour MC, Sweet KM, Xiong J, Namarika D, Mwafongo A, Nyirenda M, Chiwoko L, Kamwendo D, Hoffman I, Lee J, Phiri S, Vahrson W, Damania B, Dittmer DP. 2014. Viral profiling identifies multiple subtypes of Kaposi's sarcoma. *MBio* 5:e01633-14. <https://doi.org/10.1128/mBio.01633-14>
85. Jacobson MA, Dittmer DP, Sinclair E, Martin JN, Deeks SG, Hunt P, Mocarski ES, Shiboski C. 2009. Human herpesvirus replication and abnormal CD8+ T cell activation and low CD4+ T cell counts in antiretroviral-suppressed HIV-infected patients. *PLoS ONE* 4:e5277. <https://doi.org/10.1371/journal.pone.0005277>
86. Koelle DM, Huang ML, Chandran B, Vieira J, Piepkorn M, Corey L. 1997. Frequent detection of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) DNA in saliva of human immunodeficiency virus-infected men: clinical and immunologic correlates. *J Infect Dis* 176:94–102. <https://doi.org/10.1086/514045>
87. Futuyma DJ. 1998. *Evolutionary biology*. 3rd ed. Sinauer Associates, Sunderland, Mass.
88. Guellil M, van Dorp L, Inskip SA, Dittmar JM, Saag L, Tambets K, Hui R, Rose A, D'Atanasio E, Kriiska A, Varul L, Koekkeloren AMHC, Goldina RD, Cessford C, Solnik A, Metspalu M, Krause J, Herbig A, Robb JE, Houldcroft CJ, Scheib CL. 2022. Ancient herpes simplex 1 genomes reveal recent viral structure in Eurasia. *Sci Adv* 8:eabo4435. <https://doi.org/10.1126/sciadv.abo4435>
89. Bowen CD, Paavilainen H, Renner DW, Palomäki J, Lehtinen J, Vuorinen T, Norberg P, Hukkanen V, Szpara ML. 2019. Comparison of herpes simplex virus 1 strains circulating in finland demonstrates the uncoupling of whole-genome relatedness and phenotypic outcomes of viral infection. *J Virol* 93:e01824-18. <https://doi.org/10.1128/JVI.01824-18>
90. Liu Z, Fang Q, Zhou S, Minhas V, Wood C, He N, Zhang T. 2017. Seroprevalence of Kaposi's sarcoma-associated herpesvirus among HIV-infected uygurs in xinjiang, China. *J Med Virol* 89:1629–1635. <https://doi.org/10.1002/jmv.24804>
91. Hahn AS, Kaufmann JK, Wies E, Naschberger E, Panteleev-Ivlev J, Schmidt K, Holzer A, Schmidt M, Chen J, König S, Ensser A, Myoung J, Brockmeyer NH, Stürzl M, Fleckenstein B, Neipel F. 2012. The ephrin receptor tyrosine kinase A2 is a cellular receptor for Kaposi's sarcoma-associated herpesvirus. *Nat Med* 18:961–966. <https://doi.org/10.1038/nm.2805>
92. Dutta D, Chakraborty S, Bandyopadhyay C, Valiya Veettil M, Ansari MA, Singh VV, Chandran B. 2013. EphrinA2 regulates clathrin mediated KSHV endocytosis in fibroblast cells by coordinating integrin-associated signaling and c-cbl directed polyubiquitination. *PLoS Pathog* 9:e1003510. <https://doi.org/10.1371/journal.ppat.1003510>