Open Forum Infectious Diseases

MAJOR ARTICLE

Population Genetic Structure and Human Adaptation of Kaposi Sarcoma–Associated Herpesvirus

Alessandra Mozzi,[1] Diego Forni,[1] Rachele Cagliani,[1] Cristian Molteni,[1] Mario Clerici,[2,3] and Manuela Sironi[4]

[1]Computational Biology Unit, Scientific Institute IRCCS E. Medea, Bosisio Parini, Italy, [2]Department of Physiopathology and Transplantation, University of Milan, Milan, Italy, [3]Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy, and [4]School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

Background. Kaposi sarcoma–associated herpesvirus (KSHV), the etiologic agent of Kaposi sarcoma, is human-specific and is thought to have emerged from primate-infecting gammaherpesviruses. KSHV seroprevalence shows geographic variation, being highest in sub-Saharan Africa, intermediate in the Mediterranean area, and low in most other locations. However, KSHV prevalence is also particularly high in specific regions such as the Miyako Islands (Japan).

Methods. We retrieved KSHV genomes from public repositories and analyzed geographic patterns using principal component analysis and STRUCTURE. Adaptation to the human host was investigated by likelihood ratio tests for positive selection. Protein structures were derived from the HerpesFolds database.

Results. Most non-African genomes are genetically separated by the African genomes, and the latter are divided into 2 main lineages. The African genomes received most of their ancestry from 2 populations showing limited drift, suggesting an African origin for circulating KSHV strains. Several non-African genomes instead have most of their ancestry covered by a highly drifted ancestral population. However, some non-African genomes show similar ancestry proportions to the African ones, including those from Miyako Islands and the variant F subtype sampled in France. Molecular analysis of adaptation to the human host identified core genes as the major selection targets, including 2 viral enzymes that counteract human immune defenses.

Conclusions. We suggest that the genetic diversity of extant strains reflects relatively recent demographic events associated with viral lineage extinctions, which may have influenced KSHV epidemiology. Adaptation to the human host involved changes in core genes, possibly a strategy to optimize protein–protein interactions.

Keywords. Kaposi sarcoma–associated herpesvirus; KSHV; population genetic structure; positive selection; human adaptation.

Human herpesvirus 8 (HHV-8, order Herpesvirales, family Orthoherpesvirinae, genus *Rhadinovirus*), also known as Kaposi sarcoma–associated herpesvirus (KSHV), is the etiological agent of Kaposi sarcoma (KS), primary effusion lymphoma, multicentric Castleman disease, and KSHV inflammatory cytokine syndrome [1]. Although the virus is causally associated with tumor formation, only a small minority of infected subjects eventually develop disease. In fact, most KS cases occur in people with human immunodeficiency virus (HIV) [1]. In people with HIV, KS commonly develops upon progression to AIDS. However, it can occur even in individuals with high $CD4^+$ T-cell counts [2–4] and during antiretroviral drug treatment [5]. Moreover, HHV-8–associated malignancies can sporadically occur in HIV-seronegative individuals [6–8]. KS was first described by Moritz Kaposi in 1872, well ahead of the HIV pandemic [9]. At that time, KS was considered an uncommon, slow-growing malignancy affecting middle-aged and elderly men [9]. Until the early 1980s, KS remained a rare disease, mainly reported in Central Africa. The AIDS pandemic raised the risk of KS worldwide, giving raise to the so-called "epidemic KS," as opposed to the previously rare "endemic KS" [8]. For unknown reasons, the seroprevalence of KSHV shows important geographic variation: It reaches >90% in rural areas of sub-Saharan Africa, whereas it amounts to 20%–30% in the Mediterranean area and drops below 10% in Northern Europe, Asia, and the United States (US) [8]. However, regional differences also exist, and KSHV prevalence was reported to be particularly high in specific regions such as the Xinjiang autonomous region in China and the Miyako Islands (Okinawa Prefecture, Japan) [10, 11].

Like all other herpesviruses, HHV-8 is a double-stranded DNA virus with a long genome (~165 kb) consisting of a single unique region flanked by terminal repeats at both termini of the genome. Its life cycle is characterized by latent and lytic phases; the virus is primarily transmitted by intermittent shedding in saliva [12–14] and it enters latency upon cell infection. Systemic viral reactivation from the latent reservoir, primarily B cells, results in KSHV-related disease.

Received 18 December 2024; editorial decision 18 February 2025; accepted 20 February 2025; published online 24 February 2025

Correspondence: Alessandra Mozzi, PhD, Computational Biology Unit, Scientific Institute IRCCS E. Medea, Bosisio Parini 23842, Italy (alessandra.mozzi@lanostrafamiglia.it).

Open Forum Infectious Diseases®

https://doi.org/10.1093/ofid/ofaf111

Another feature that KSHV shares with other herpesviruses is its extremely limited natural host range, which is restricted to humans [15]. This is generally considered the result of very long-standing adaptation and co-speciation of these viruses with their hosts. Indeed, viruses phylogenetically related to KSHV were isolated from nonhuman primates. Specifically, in Old World primates of the genera *Macaca* and *Colobus*, 2 rhadinovirus lineages were identified: one more closely related to KSHV, which includes the retroperitoneal fibromatosis–associated herpesvirus and colobine gammaherpesvirus 1, and a second one that contains rhesus macaque rhadinovirus, Macaca nemestrina rhadinovirus, and Japanese macaque rhadinovirus [16]. Rhadinoviruses related to KSHV were also detected in chimpanzees and gorillas, but their full genome sequences have not been obtained [17, 18]. Finally, distinct rhadinoviruses, with lower similarity to KSHV, were detected in New World primates [19].

MATERIALS AND METHODS

Viral Genome Sequences, Alignment, and Network

Viral genome sequences were selected from the Bacterial and Viral Bioinformatics Resource Center portal (http://www.bv-brc.org/). A detailed list of accession numbers is reported in Supplementary Table 1. A phylogenetic network was generated using SplitsTree4 (v4.16.2) [20]. Genomic alignments were generated using MAFFT with default parameters [21]. Alignments were uploaded in SeaView [22] and visually inspected. Biallelic parsimony-informative (PI) sites with a minimum frequency of 2 and with 90% of sequence coverage were selected. A principal component analysis (PCA) was performed using the PI matrix as input.

T-distributed stochastic neighbor embedding (tSNE) was applied to visualize the high-dimensional single-nucleotide polymorphism (SNP) data in a 2-dimensional plot. tSNE was calculated with the R package Rtsne [23], with an input matrix composed by SNPs covered by all viral strains (n = 1712), setting the perplexity parameter at 30 and run for 5000 iterations.

Linkage Disequilibrium

To evaluate linkage disequilibrium (LD), we used LIAN software (v.3.7) [24]. Significance was assessed by Monte Carlo simulations (1000 iterations).

Population STRUCTURE Analysis

The PI data used for PCA analysis were also used to run STRUCTURE [25, 26]. The optimal K was evaluated with the Evanno method [27] using the HARVESTER tool [28]. The amount of drift that each subpopulation experienced from a hypothetical ancestral population was quantified by the F parameter calculated for the optimal K value (K = 4) [25].

Molecular Dating

The largest sequences (~1227 bp, *ORF26*) available for a sizable number of strains sampled before 1995 (Supplementary Table 2) was used for root-to-tip regression. A maximum-likelihood tree was generated by IQTREE v1.6.12 [29]. We then calculated the correlation coefficient (r) of the regression of root-to-tip genetic distances against sequence sampling years.

Detection of Positive Selection in the HHV-8 Lineage

We analyzed a viral phylogeny composed of 12 different strains (Supplementary Table 3): 5 fully sequenced viruses infecting Old World monkey species and 7 HHV-8 strains selected from the PCA analysis.

For each viral genome, we retrieved coding sequences of all annotated open reading frames (ORFs) (Supplementary Table 4). Gene alignments were generated using GUIDANCE2 [30]. For each coding gene, phylogenetic trees were reconstructed using phyML. Episodic positive selection on the HHV-8 branch was detected by applying the branch-site likelihood ratio tests from codeml ("test 2") [31]. To identify sites evolving under positive selection, we used Bayes Empirical Bayes analysis.

Structural Mapping of Positively Selected Sites

Three-dimensional protein structure available for capsid vertex component 2 (ORF19, PDB ID: 7nxq) was obtained from the Protein Data Bank archive (www.rcsb.org, last accessed 29 July 2024). The molecular structures of all other proteins were generated from the HerpesFolds database (https://www.herpesfolds.org/herpesfolds) [32].

A detailed version of this section is available in the Supplementary Methods.

RESULTS

Genetic Diversity and Geographic Associations of Circulating KSHV Strains

We obtained all available complete or almost complete KSHV genomes from public databases (Supplementary Table 1). Sequences with known country of origin (n = 164) were aligned and a neighbor-net split network was generated (Figure 1*A*). Despite extensive reticulation, suggestive of recombination, some clusters were evident and showed association with geographic areas. Whereas some of such clusters mostly included sequences sampled in Africa, many non-African genomes defined a single cluster with apparently limited genetic diversity (Figure 1*A*).

To gain insight into the genetic relatedness of KSHV genomes, we applied PCA (Figure 1*B*). The first principal component (PC), which explained 21% of the variance, mainly separated sequences sampled in sub-Saharan Africa. Some evidence of clustering in this geographic area was observed, with west-to-east and north-to-south gradients. In fact, sequences from West Africa (Cameroon) tended to separate from those sampled in Central Africa (Uganda) and Southern Africa

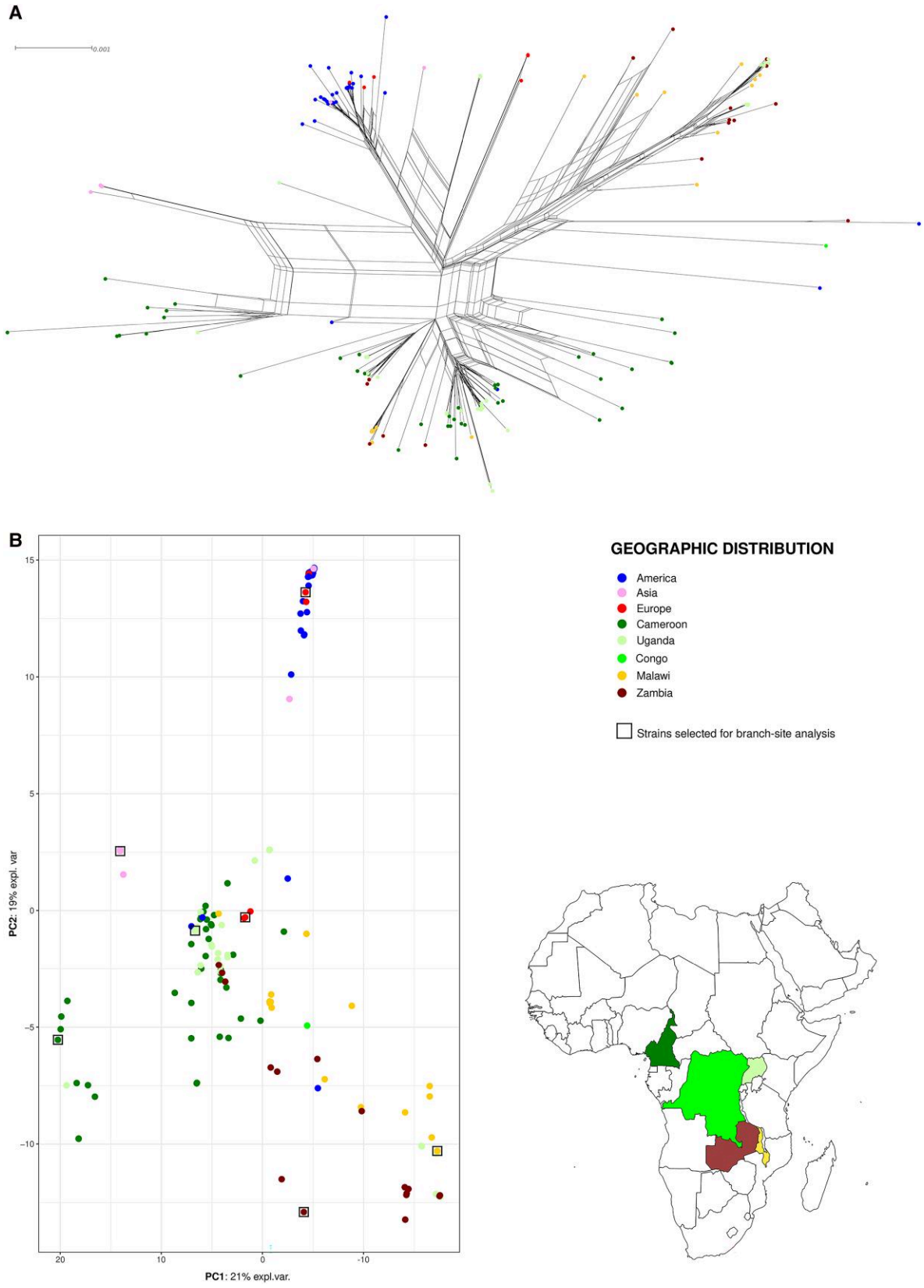

Figure 1. Divergence among human herpesvirus 8 (HHV-8) strains. *A*, Neighbor-net split network of HHV-8 genomes (n = 164). Each sequence is shown as a dot, color-coded according to geographical location, as described in the key. *B*, Principal component (PC) analysis of HHV-8 whole-genome diversity. Onto the Cartesian axes (y and x), the first and second components are reported, which describe 21% and 19% of genome diversity. Each dot is a genome, color-coded as in *A*. On the right, a geographic map of Africa shows countries color-coded according to geographical information of African HHV-8 isolates.

(Malawi and Zambia). However, the clustering was clearly incomplete, with sequences from Uganda having the widest distribution along the first PC. The second PC (19% of variance) clearly placed most non-African sequences in a distinct cluster, but also contributed to the separation of the African ones. Seven genomes from France, Canada, and the US instead remained in the African cluster. Also, 4 of the 5 Japanese sequences (including 3 from Miyako Islands in Okinawa Prefecture) formed a distinct cluster. For further validation, analyses were also performed using a different approach, namely tSNE. Results very similar to those in the PCA were obtained, with the formation of 3 major clusters comprising non-African sequences (plus a single sequence from Uganda), mixed genomes, and Africa-only strains (Supplementary Figure 1).

Analysis of KSHV Populations

To gain further insight into the structure of KSHV populations, we used the program STRUCTURE, which relies on a Bayesian statistical model for clustering genotypes into populations without prior information on their genetic relatedness [25, 26, 33]. The program can identify distinct subpopulations (or clusters [K]) that compose the overall population. Subpopulations can then be related to specific features such as geographic origin or genotype classification. Because STRUCTURE is ideally suited for weakly linked markers [25], we first analyzed the level of LD with LIAN v3.7 [24]. Statistically significant LD was detected ($P < .01$), with a standardized index of association of 0.0692, a value indicating very weak LD and allowing application of STRUCTURE models.

We used the linkage model with correlated allele frequencies, which assumes that discrete genome "chunks" were inherited from K ancestral populations [25]. To estimate the optimal number of subpopulations, we ran STRUCTURE for values of K from 1 to 14. The ΔK method yielded a major peak at K = 4 (Supplementary Figure 2). Analysis of ancestry components was thus performed for 4 subpopulations (Figure 2*A*). For each of these subpopulations, we used the linkage model in STRUCTURE to estimate the F parameter, which represents a measure of genetic differentiation between populations based on allele frequencies (Figure 2*B*). Results indicated that the lowest drift was experienced by 1 subpopulation (Africa_1) that accounted for the largest ancestry component of all Cameroonian samples, as well as most sequences from Uganda. The Africa_1 population also contributed considerable ancestry proportions to some sequences from Zambia and Malawi. This population also appears at high proportions in a few non-African samples, which correspond to some of those that failed to separate on the second PC. These include the F variant strains, which were associated with severe clinical presentation in French patients, whereas the reference F strain was derived from a Congolese patient with multicentric Castleman disease [34]. The 2 populations with intermediate drift (Africa_2 and Worldwide) had very different distributions. Africa_2 accounted for a major ancestry component of several genomes sampled in Zambia, Uganda, and Malawi, being poorly represented elsewhere. The worldwide component instead occurred with variable proportions in most sequences. Finally, the most drifted component (non-Africa) was observed at very high frequency in most non-African genomes and only accounted for minor ancestry components of African sequences. Compared to other non-African strains, sequences from Japan, including those from Okinawa Prefecture, had lower representation of the non-African component and more of the Worldwide and Africa_1 components (Figure 2).

Overall, these data are consistent with the presence of 2 major KSHV lineages in Africa [35], which are characterized by the Africa_1 and Africa_2 components. Most non-African genomes have instead their ancestry contributed by a highly drifted ancestral population with relatively low representation in Africa.

Molecular Dating of KSHV

Next, we aimed to estimate the time of KSHV emergence using tip dating. To this purpose, a heterochronous dataset is necessary. We thus searched public databases for KSHV genome fragments sampled across the widest possible time interval. We identified 42 sequences of ORF26 that were sampled from the 1970s to the 1990s, and we combined them with the ORF26 sequences from our full genome dataset. We thus obtained an alignment of 177 sequences with sample dates between 1971 and 2021 (Supplementary Table 2). Reliable time estimates can only be obtained if a temporal signal is detected in the sampled sequences. Thus, we performed regression of root-to-tip genetic distances against sampling dates. No significant correlation was found ($r = -0.03$), indicating that there is no temporal signal in the data and that molecular dating cannot be pursued (Supplementary Figure 3).

Genome-wide Scan of Positive Selection on the KSHV Branch

As mentioned above, KSHV is a human-specific pathogen, implying that the virus must have evolved to infect our species. We thus sought to identify the molecular mechanisms that contributed to human adaptation. To this aim, we exploited the availability of complete genomic sequences of rhadinoviruses that infect other primates. Specifically, we selected a set of genomes from rhadinoviruses that infect macaques and colobes, as well as 7 KSHV strains representative of the genetic diversity of circulating strains, according to the PCA analysis (Figure 1*B*, Supplementary Table 3). To detect adaptation to the human host, we applied a branch-site test [31] and we designated as foreground branches those that lead to the HHV-8 sequences (Figure 3). The test was performed for 66 coding genes, for which we confidently retrieved orthologous sequences from nonhuman primate rhadinoviruses (Supplementary Table 4).

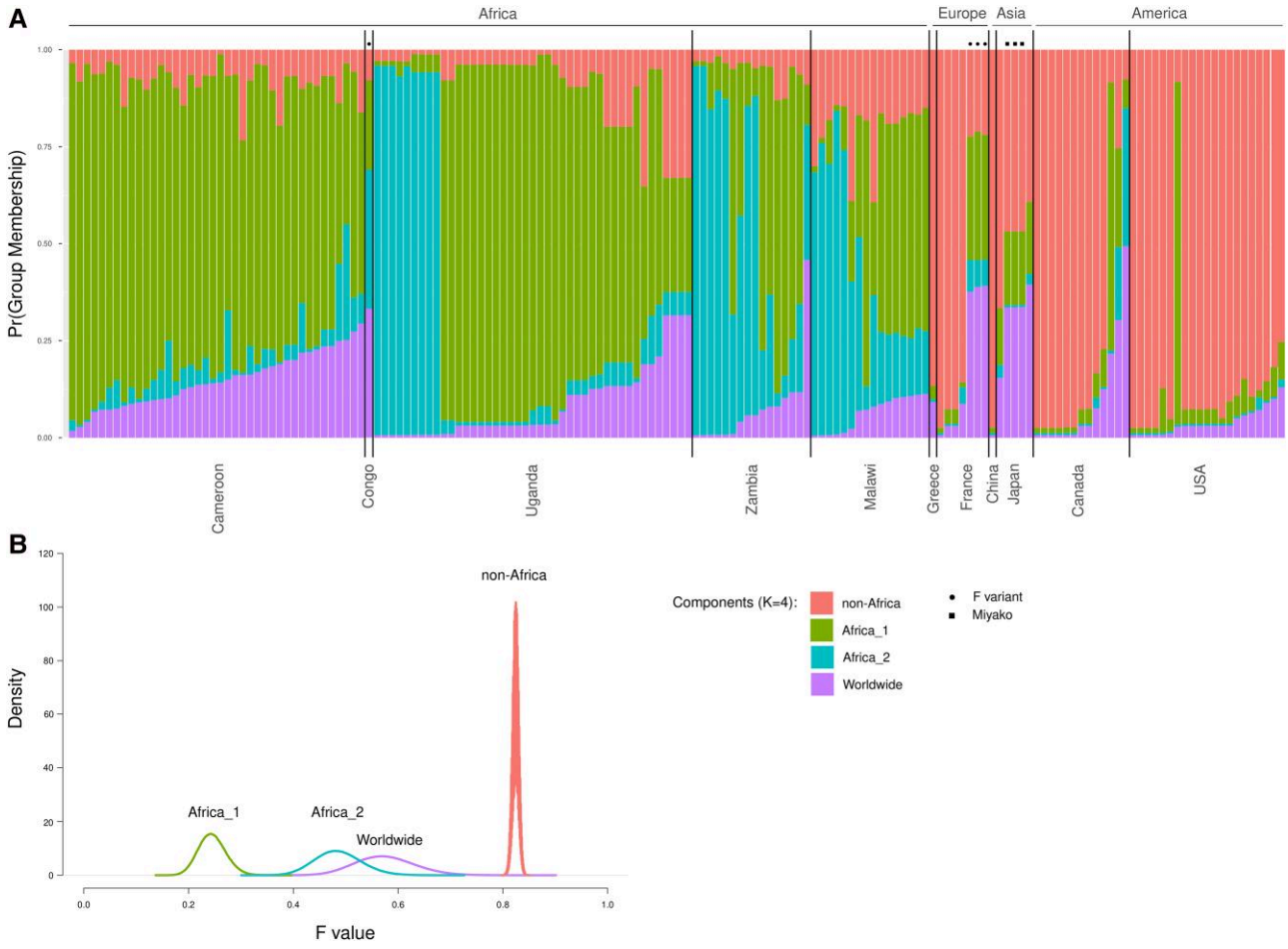

Figure 2. Population structure analysis of human herpesvirus 8 (HHV-8) strains. *A*, Bar plot representing the proportion of ancestral population components for K = 4. Each HHV-8 genome is represented as a vertical line colored according to the proportion of sites that have been assigned to the 4 ancestral components by STRUCTURE. Genomes are grouped according to sampling location. *B*, Distributions of F values for the 4 populations. Ancestry components are named based on the geographical region where they are more prevalent. Colors are as shown in *A*. Abbreviation: Pr, probability.

After accounting for recombination, we found evidence of positive selection in 6 genes: ORF19, ORF33, ORF36, ORF39, ORF42, and ORF54 (Table 1). All of these are core genes shared among all herpesviruses [36].

The branch-site test also provides information on which codons were targeted by selection. Positively selected sites were found in all genes, except for ORF39 (glycoprotein M) (Table 1). To analyze the location of the selected sites relative to protein structural domain, we searched for 3-dimensional structures of KSHV proteins. The crystal protein structure was only solved for capsid vertex component 2 (ORF19). For proteins encoded by ORF33, ORF36, ORF42, and ORF54 we used the 3-dimensional models available from the HerpesFolds database [32].

Mapping of the positively selected sites onto the molecular models, as well as on the structure of the capsid vertex component 2, indicated that most of them are surface exposed (Figure 4). The capsid vertex component 2 protein is a key component of the viral capsid. The 2 positively selected sites are located in the globular domain of the protein. Although not directly involved in the interaction with other pORF19 monomers, these 2 sites can potentially interact with other proteins, as ORF19 is also a component of the capsid-associated tegument complexes [37].

Signals of positive selection were also found in 2 cytoplasmic envelopment proteins (cep1:ORF42 and cep2:ORF33), which are involved in tegumentation and envelope acquisition within the host cytoplasm, playing a direct role in viral egress [38, 39] (Figure 4). These 2 proteins are predicted to interact with many other viral proteins. The interaction of ORF33 from MHV-68 (murine gammaherpesvirus 68, which is closely related) with ORF25 and ORF26 capsid proteins and with ORF38 and ORF45 tegumental proteins was experimentally validated, but the molecular determinants of these interactions remain still unknown. Only the ORF33 N-terminal domain (1–17 aa) was demonstrated to be essential for interaction with ORF45

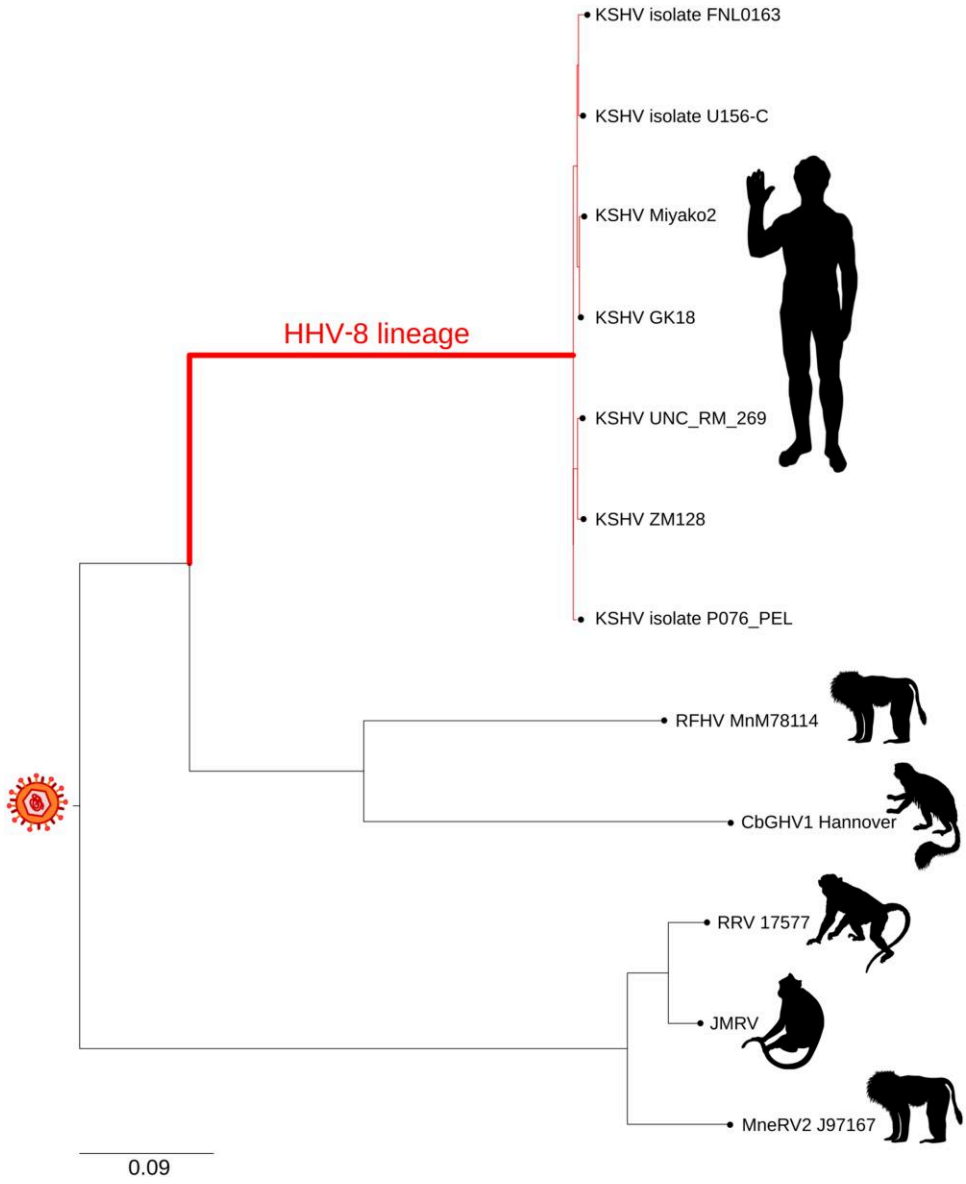

Figure 3. Branch-site analysis. A maximum-likelihood tree of *ORF9* (encoding the DNA polymerase catalytic subunit) is drawn to exemplify the phylogenetic relationships among primate rhadinoviruses (strain information and GenBank IDs are reported in Supplementary Table 3). The tree was constructed using PhyML and visualized using FigTree. The human herpesvirus 8 branch, which was specifically tested for episodic positive selection, is shown in red. Silhouette images were retrieved from PhyloPic (www.phylopic.org). Abbreviations: CbGHV1, colobine gammaherpesvirus 1; HHV-8, human herpesvirus 8; JMRV, Japanese macaque rhadinovirus; KSHV, Kaposi sarcoma–associated herpesvirus; MneRV2, Macaca nemestrina rhadinovirus; RFHV, retroperitoneal fibromatosis–associated herpesvirus; RRV, rhesus macaque rhadinovirus.

(yellow, Figure 4) [39]. Interestingly, 2 positively selected sites (V20 and G25) are just downstream from this region, and together with the P93 residue define a restricted protein edge, opposed to the alpha helix involved in ORF45 recruitment (Figure 4). ORF42 is also considered a tegumental protein, but it was recently identified to have a peculiar function in promoting viral protein expression, a feature that may be restricted to HHV-8 [40]. Recently, ORF42 was shown to interact with ORF55, and a variant typical of strains from Okinawa Prefecture (R136Q) was found to affect interaction with ORF55 and mature virion production. Mapping of the R136 position indicates that it is located in the same surface area as the positively selected C69 site [41].

Finally, ORF36 and ORF54 encode 2 viral enzymes: the viral kinase (vPK) and the deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase). In vPK, positively selected sites are not located in the region directly involved in ATP binding or in the structures involved in the binding of the RNA helicase [42, 43]. In dUTPase, most of the positively selected sites are in the C-terminal part of the protein and 4 of them fall within the dUTPase-related domain, which was suggested to have evolved to enable interactions with other proteins [44].

Table 1. Likelihood Ratio Test Statistics for Models of Variable Selective Pressure on the Kaposi Sarcoma–Associated Herpesvirus (Human Herpesvirus 8) Branch

		MA vs MA1		
Gene	Protein Name	*P* Value[a]	−Δ2lnL[b]	BEB Sites[c]
ORF19	Capsid vertex component 2	.0437	8.30	P271, S344
ORF33	Cytoplasmic envelopment protein 2	.0212	10.26	V20, G25, P93, C152, G178, S269
ORF36	Viral protein kinase	.0212	9.95	N298, G332
ORF39	Envelope glycoprotein M	.0180	11.20	…
ORF42	Cytoplasmic envelopment protein 1	.0180	11.92	C69, S225
ORF54	Deoxyuridine 5′-triphosphate nucleotidohydrolase	.0002	21.57	T51, P120, N134, E169, R226, F233, H254

Abbreviations: BEB, Bayes Empirical Bayes.

[a]False discovery rate–corrected *P* values.

[b]2ΔlnL: twice the difference of the natural logs of the maximum likelihood of the models being compared.

[c]Positions refer to proteins of human herpesvirus 8 strain GK18 (NC_009333). Sites had a posterior probability >0.95 to have $\omega > 1$.

Interestingly, the selected sites define a specific linear surface–exposed patch (Figure 4).

DISCUSSION

The seroprevalence of KSHV shows remarkable geographic variation [8]. The high prevalence observed in sub-Saharan Africa was suggested to be due to environmental factors, such as coinfection with malaria and other parasites, or behavioral practices that expose children to saliva [45, 46]. However, prevalence varies even outside Africa, being higher in the Mediterranean regions, for instance, than in several other locations. Moreover, some specific geographic areas are known to have an extremely high prevalence of KSHV, including the Xinjiang region of China [47], the Miyako Islands (Okinawa Prefecture, Japan) [11], and Peru [48]. Factors such as historical migrations and human genetic diversity were proposed as explanations, but definitive causal links are missing [11, 49]. Our study was thus motivated by the idea that a better understanding of KSHV diversity and evolution might provide further insight into its epidemiology.

Confirming previous results [35, 50], our analyses showed that most non-African genomes are genetically separated by the African ones and that the latter are divided into 2 main lineages. Population structure analysis indicated that the African genomes received most of their ancestry from 2 populations showing less drift than the worldwide and non-African populations. Specifically, population Africa_1, which is most common in genomes from Cameroon and Uganda, experienced the lowest drift. Overall, these data strongly suggest an African origin for circulating KSHV strains, with a possible emergence in a central region followed by diversification and southward migration. Several non-African strains instead have most of their ancestry covered by a highly drifted ancestral population. This is in principle consistent with the idea that the out-of-Africa event was accompanied by a bottleneck with consequent genetic drift. It is thus tempting to conclude that KSHV infected anatomically modern humans as they evolved in Africa and migrated from the continent with its hosts, an event that affected the genetic diversity of both the virus and of human populations [51]. In this scenario, the non-African genomes that show similar ancestry proportions to the African ones may represent instances of recent migrations. However such strains also include the 3 from Miyako Islanders, which the PCA also identified to be most closely related to African strains compared to most other non-African genomes. Given the high prevalence of the virus in these islands and their remote geographic position, it is unlikely that the 3 sequences derive from recent migrations. One alternative possibility is that the genetic diversity of extant strains reflects more recent events associated with viral lineage extinctions, as previously shown in the case of other herpesviruses [52, 53, 54]. As an example, circulating strains of herpes simplex virus (HSV) type 1 were shown to have originated 7000 to 5000 years ago [53, 54], well after the initial out-of-Africa migration of modern humans (dated around 60 000 years ago) [51]. This is not surprising, as following the initial worldwide dispersal, human migration patterns across continents often resulted in the replacement of existing populations. For instance, large-scale population migrations and replacements during the Bronze Age occurred in Eurasia [51, 55]. Turnover in human populations might have caused the consequential replacement of KSHV strains. If this were the case, viruses with a strong component of ancestry contributed by the non-African population might have replaced in several areas those circulating previously, whereas other regions (including the Miyako Islands) may have instead retained the original strains. Unfortunately, these hypotheses are impossible to test, as molecular dating of KSHV evolutionary history is presently not applicable. Extensive geographic sampling of the diversity of KSHV or the sequencing of viral genomes from archaeological remains might in the future allow dating of the origin and out-of-Africa spread of KSHV. Alternatively, new methods that incorporate modeling of the decrease of substitution rates with the timescale of measurement may provide insight into KSHV evolutionary history [56, 57]. Based on present data, it is however worth noting that the viruses sampled in a high-

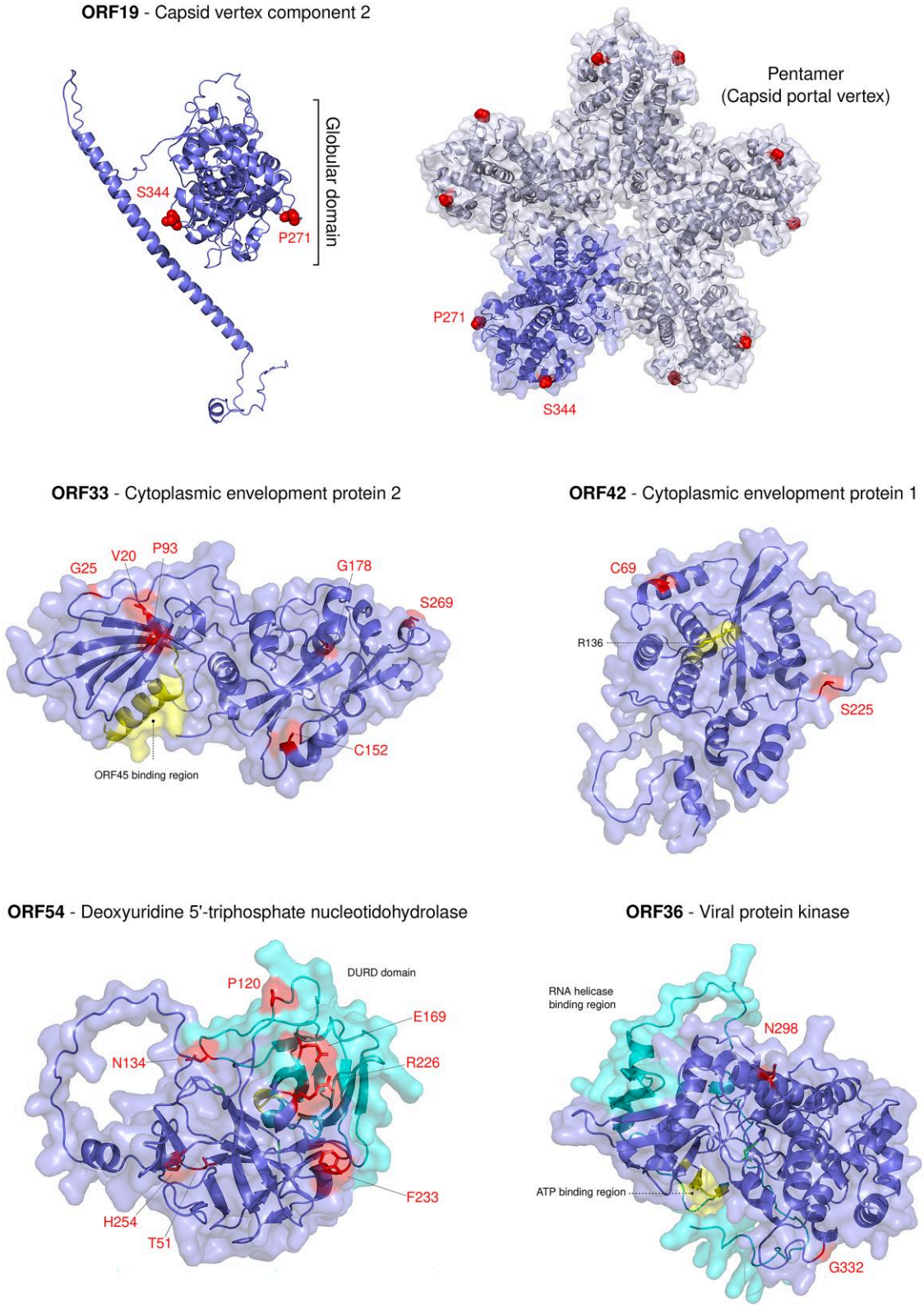

Figure 4. Positive selection in human herpesvirus 8 lineages. Three-dimensional protein structure models of ORF 33, ORF42, ORF54, and ORF36 are reported as illustrations contoured by protein surface. Positively selected sites are shown in red. For ORF19, positively selected sites were mapped onto the pentameric structure of the capsid portal vertex, solved by X-ray diffraction (Protein Data Bank ID: 7nxq). Yellow highlighting indicates binding region for ORF45 onto the ORF33 model, R136 site onto the ORF42 model, and adenosine triphosphate (ATP)–binding region onto the ORF36 model. The extended deoxyuridine 5′-triphosphate nucleotidohydrolase–related domain (DURD) of ORF54 and the large region of ORF36 involved in RNA helicase binding are shown in cyan.

prevalence region, the Miyako Islands, have a different genetic history than those that are common outside Africa. Whether this affects some virus phenotypes associated with transmissibility and, as a consequence, seroprevalence remains to be evaluated.

Another interesting observation concerns the viruses sampled in France, Canada, and the US that have similar genetic ancestry as those sampled in Africa. Whereas it is impossible to determine if they derive from historical or recent migrations, in France these viruses belong to the variant F subtype, which was recently described in a study that investigated KSHV-related disease in men who have sex with men (MSM) [34]. The variant F subtype was associated with a severe clinical presentation in the context of immunosuppression and was found to circulate in a restricted MSM population [34]. It was thus suggested that this new variant has a recent origin and has only recently spread in humans [58]. Our data do not support this view, as the ancestry components of these viruses are similar to some sampled in Africa and are not particularly drifted.

Irrespective of the origin and evolutionary history of extant strains, it is generally considered that human herpesviruses emerged from primate viruses through cospeciation or host switching [59, 60]. Because these viruses are highly species-specific in nature, they must have adapted to infect their hosts, and the changes that contributed to such adaptation should be traceable in their phylogenetic trees. We thus used a phylogeny of primate rhadinoviruses to search for evidence of positive selection on the HHV-8 branch. The positively selected genes we detected are all core genes. One of them encodes the capsid vertex component 2 (pORF19), which is essential for capsid assembly and virion release [61], whereas other 2 (ORF42 and ORF33; cytoplasmic envelopment proteins 1 and 2, respectively) are involved in tegumentation and envelope acquisition within the host cytoplasm, playing a direct role in viral egress [38, 39]. Interestingly, the HSV type 2 (HSV-2) ortholog (UL25) of KSHV ORF19 was also found to contribute to adaptation to the human host, and the same holds true for other capsid components encoded by human cytomegalovirus (HCMV) [62, 63]. More generally, similar to our findings in KSHV, proteins involved in virion formation and egress were targeted by selection in both HSV-2 and HCMV and contributed to host adaptation [62, 63]. One possible explanation for these common findings is that, because virion maturation in herpesviruses is orchestrated by the interplay of viral and host proteins [64], the positive selection signals derive from optimization of protein–protein interactions.

Selection in 2 viral enzymes, vPK (ORF36) and dUTPase (ORF54), was observed as well. These proteins play a number of functions in infected cells, which are related or unrelated to their enzymatic activity. Interestingly, both proteins were shown to counteract the host immune defenses. In fact, vPK inhibits IRF-3–mediated type I interferon responses [65], and ORF54 mediates the degradation of IFNAR1 [66], as well as the downregulation of the NKp44 ligands, which are involved in the activation of natural killer cells [67]. It is thus likely that the selective pressure on these viral genes was exerted by the necessity of modulating the host immune response to the advantage of the virus.

Our study clearly has limitations. First, the available number of fully sequenced KSHV genomes is relatively small and their geographic distribution uneven. Most sequences were sampled in Africa, and several high-prevalence areas (eg, Xinjiang and Peru) are not represented. This possibly introduces biases and limits our ability to relate evolutionary history and epidemiology. Biases are also likely to be caused by the fact that most African sequences derive from patients with KS, whereas the genomes associated to KSHV inflammatory cytokine syndrome or primary effusion lymphoma were sampled outside Africa. Thus, additional sequencing of KSHV genomes will be instrumental to provide insight into the evolution of the viral populations, as well as into disease epidemiology and clinical presentation.

Supplementary Data

Supplementary materials are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. This work was supported by the Italian Ministry of Health ("Ricerca Corrente").

Author contributions. Conceptualization: M. S. and A. M. Formal analysis: A. M., D. F., C. M., and R. C. Investigation: M. S., A. M., and D. F. Visualization: A. M. and C. M. Writing—original draft: M. S. and A. M. Writing—review and editing: R. C., D. F., M. C., and M. S. Funding acquisition: M. S. All authors have read and agreed to the published version of the manuscript.

Data availability. *Data are available in the Supplementary Material.*

Potential conflicts of interest. All authors: No reported conflicts of interest.

References

1. Dittmer DP, Damania B. Kaposi sarcoma–associated herpesvirus: immunobiology, oncogenesis, and therapy. J Clin Invest **2016**; 126:3165–75.
2. Maurer T, Ponte M, Leslie K. HIV-associated Kaposi's sarcoma with a high CD4 count and a low viral load. N Engl J Med **2007**; 357:1352–3.
3. Krown SE, Lee JY, Dittmer DP. More on HIV-associated Kaposi's sarcoma. N Engl J Med **2008**; 358:535–6.
4. Von Braun A, Braun DL, Kamarachev J, Günthard HF. New onset of Kaposi sarcoma in a human immunodeficiency virus-1–infected homosexual man, despite early antiretroviral treatment, sustained viral suppression, and immune restoration. Open Forum Infect Dis **2014**; 1:ofu005.
5. Shiels MS, Engels EA. Evolving epidemiology of HIV-associated malignancies. Curr Opin HIV AIDS **2017**; 12:6–11.
6. Vangipuram R, Tyring SK. Epidemiology of Kaposi sarcoma: review and description of the nonepidemic variant. Int J Dermatol **2019**; 58:538–42.
7. Ziegler J, Newton R, Bourboulia D, et al. Risk factors for Kaposi's sarcoma: a case-control study of HIV-seronegative people in Uganda. Int J Cancer **2003**; 103:233–40.
8. Cesarman E, Damania B, Krown SE, Martin J, Bower M, Whitby D. Kaposi sarcoma. Nat Rev Dis Primers **2019**; 5:9.
9. Sternbach G, Varon J. Moritz Kaposi: idiopathic pigmented sarcoma of the skin. J Emerg Med **1995**; 13:671–4.
10. Zhang T, Shao X, Chen Y, et al. Human herpesvirus 8 seroprevalence, China. Emerg Infect Dis **2012**; 18:150–2.
11. 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**; 216:850–8.

12. Koelle DM, Huang ML, Chandran B, Vieira J, Piepkorn M, Corey L. 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 **1997**; 176:94–102.
13. Pauk J, Huang M-L, Brodie SJ, et al. Mucosal shedding of human herpesvirus 8 in men. N Engl J Med **2000**; 343:1369–77.
14. Casper C, Krantz E, Selke S, et al. Frequent and asymptomatic oropharyngeal shedding of human herpesvirus 8 among immunocompetent men. J Infect Dis **2007**; 195:30–6.
15. Azab W, Dayaram A, Greenwood AD, Osterrieder N. How host specific are herpesviruses? Lessons from herpesviruses infecting wild and endangered mammals. Annu Rev Virol **2018**; 5:53–68.
16. Dhingra A, Ganzenmueller T, Hage E, et al. Novel virus related to Kaposi's sarcoma–associated herpesvirus from colobus monkey. Emerg Infect Dis **2019**; 25:1548–51.
17. Lacoste V, Mauclère P, Dubreuil G, Lewis J, Georges-Courbot M-C, Gessain A. A novel γ2-herpesvirus of the rhadinovirus 2 lineage in chimpanzees. Genome Res **2001**; 11:1511–9.
18. Lacoste V, Mauclre P, Dubreuil G, Lewis J, Georges-Courbot M-C, Gessain A. KSHV-like herpesviruses in chimps and gorillas. Nature **2000**; 407:151–2.
19. Fickenscher H, Fleckenstein B. Herpesvirus saimiri. Phil Trans R Soc Lond B **2001**; 356:545–67.
20. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **2006**; 23:254–67.
21. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **2013**; 30:772–80.
22. Galtier N, Gouy M, Gautier C. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci **1996**; 12:543–8.
23. Van Der Maaten L. Accelerating t-SNE using tree-based algorithms. J Mach Learn Res **2014**; 15:3221–45.
24. Haubold B, Hudson RR. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage analysis. Bioinformatics **2000**; 16:847–8.
25. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **2003**; 164:1567–87.
26. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics **2000**; 155:945–59.
27. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol **2005**; 14:2611–20.
28. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour **2012**; 4:359–61.
29. Trifinopoulos J, Nguyen LT, Haeseler Av, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res **2016**; 44(W1):232–5.
30. Sela I, Ashkenazy H, Katoh K, Pupko T. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res **2015**; 43(W1):7–14.
31. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol **2005**; 22:2472–9.
32. Soh TK, Ognibene S, Sanders S, Schäper R, Kaufer BB, Bosse JB. A proteome-wide structural systems approach reveals insights into protein families of all human herpesviruses. Nat Commun **2024**; 15:10230.
33. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour **2009**; 9:1322–32.
34. 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**; 222:1320–8.
35. Moorad R, Juarez A, Landis JT, et al. Whole-genome sequencing of Kaposi sarcoma–associated herpesvirus (KSHV/HHV8) reveals evidence for two African lineages. Virology **2022**; 568:101–14.
36. Davison AJ. Comparative analysis of the genomes. In: Arvin A, Campadelli-Fiume G, Mocarski E, et al, eds. Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge, UK: Cambridge University Press, **2007**:10–26.
37. Gong D, Dai X, Jih J, et al. DNA-packing portal and capsid-associated tegument complexes in the tumor herpesvirus KSHV. Cell **2019**; 178:1329–43.e12.
38. Shen S, Jia X, Guo H, Deng H. Gammaherpesvirus tegument protein ORF33 is associated with intranuclear capsids at an early stage of the tegumentation process. J Virol **2015**; 89:5288–97.
39. Jia X, Sun L, Shen S, et al. The interaction between tegument proteins ORF33 and ORF45 plays an essential role in cytoplasmic virion maturation of a gammaherpesvirus. J Virol **2022**; 96:e01073-22.
40. Butnaru M, Gaglia MM. The Kaposi's sarcoma–associated herpesvirus protein ORF42 is required for efficient virion production and expression of viral proteins. Viruses **2019**; 11:711.
41. Kuriyama K, Watanabe T, Ohno S. Analysis of the interaction between the ORF42 and ORF55 proteins encoded by Kaposi's sarcoma–associated herpesvirus. Arch Virol **2024**; 169:98.
42. Jong JE, Park J, Kim S, Seo T. Kaposi's sarcoma–associated herpesvirus viral protein kinase interacts with RNA helicase a and regulates host gene expression. J Microbiol **2010**; 48:206–12.
43. Park J, Lee D, Choe J, Seo T, Chung J. Kaposi's sarcoma–associated herpesvirus (human herpesvirus-8) open reading frame 36 protein is a serine protein kinase. J Gen Virol **2000**; 81:1067–71.
44. Davison AJ, Stow ND. New genes from old: redeployment of dUTPase by herpesviruses. J Virol **2005**; 79:12880–92.
45. Wakeham K, Webb EL, Sebina I, et al. Parasite infection is associated with Kaposi's sarcoma associated herpesvirus (KSHV) in Ugandan women. Infect Agents Cancer **2011**; 6:15.
46. Butler LM, Neilands TB, Mosam A, Mzolo S, Martin JN. A population-based study of how children are exposed to saliva in KwaZulu-Natal province, South Africa: implications for the spread of saliva-borne pathogens to children. Trop Med Int Health **2010**; 15:442–53.
47. He F, Wang X, He B, et al. Human herpesvirus 8: seroprevalence and correlates in tumor patients from Xinjiang, China. J Med Virol **2007**; 79:161–6.
48. Mohanna S, Portillo J-A, Carriquiry G, et al. Human herpesvirus-8 in Peruvian blood donors: a population with hyperendemic disease? Clin Infect Dis **2007**; 44:558–61.
49. Liu Z, Fang Q, Zuo J, et al. Was Kaposi's sarcoma–associated herpesvirus introduced into China via the ancient Silk Road? An evolutionary perspective. Arch Virol **2017**; 162:3061–8.
50. Marshall VA, Cornejo Castro EM, Goodman CA, et al. 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 **2024**; 20:e1012338.
51. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. Nature **2017**; 541:302–10.
52. Pontremoli C, Forni D, Clerici M, Cagliani R, Sironi M. Possible European origin of circulating varicella-zoster virus strains. J Infect Dis **2020**; 221:1286–94.
53. Forni D, Pontremoli C, Clerici M, Pozzoli U, Cagliani R, Sironi M. Recent out-of-Africa migration of human herpes simplex viruses. Mol Biol Evol **2020**; 37:1259–71.
54. Guellil M, Dorp Lv, Inskip SA, et al. Ancient herpes simplex 1 genomes reveal recent viral structure in Eurasia. Sci Adv **2022**; 8:eabo4435.
55. Allentoft ME, Sikora M, Sjogren KG, et al. Population genomics of Bronze Age Eurasia. Nature **2015**; 522:167–72.
56. Ghafari M, Simmonds P, Pybus OG, Katzourakis A. A mechanistic evolutionary model explains the time-dependent pattern of substitution rates in viruses. Curr Biol **2021**; 31:4689–96.e5.
57. Simmonds P, Aiewsakun P, Katzourakis A. Prisoners of war—host adaptation and its constraints on virus evolution. Nat Reviews Microbiol **2019**; 17:321–8.
58. 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**; 222:1250–3.
59. Brito AF, Baele G, Nahata KD, Grubaugh ND, Pinney JW. Intrahost speciations and host switches played an important role in the evolution of herpesviruses. Virus Evol **2021**; 7:veab025.
60. Wertheim JO, Hostager R, Ryu D, et al. Discovery of novel herpes simplex viruses in wild gorillas, bonobos, and chimpanzees supports zoonotic origin of HSV-2. Mol Biol Evol **2021**; 38:2818–30.
61. Naniima P, Naimo E, Koch S, et al. Assembly of infectious Kaposi's sarcoma–associated herpesvirus progeny requires formation of a pORF19 pentamer. PLoS Biol **2021**; 19:e3001423.
62. Mozzi A, Forni D, Cagliani R, Clerici M, Pozzoli U, Sironi M. Intrinsically disordered regions are abundant in simplexvirus proteomes and display signatures of positive selection. Virus Evol **2020**; 6:veaa028.
63. Mozzi A, Biolatti M, Cagliani R, et al. Past and ongoing adaptation of human cytomegalovirus to its host. PLoS Pathog **2020**; 16:e1008476.
64. Tandon R, Mocarski ES. Viral and host control of cytomegalovirus maturation. Trends Microbiol **2012**; 20:392–401.
65. Hwang S, Kim KS, Flano E, et al. Conserved herpesviral kinase promotes viral persistence by inhibiting the IRF-3–mediated type I interferon response. Cell Host Microbe **2009**; 5:166–78.
66. Leang RS, Wu T-T, Hwang S, et al. The anti-interferon activity of conserved viral dUTPase ORF54 is essential for an effective MHV-68 infection. PLoS Pathog **2011**; 7:e1002292.
67. Madrid AS, Ganem D. Kaposi's sarcoma–associated herpesvirus ORF54/dUTPase downregulates a ligand for the NK activating receptor NKp44. J Virol **2012**; 86:8693–704.