CTMI (2006) 312:1–42

Modern Evolutionary History of the Human KSHV Genome

G. S. Hayward (✉) · J.-C. Zong

Viral Oncology Program, The Sidney Kimmel Comprehensive Cancer Center, The Johns Hopkins University School of Medicine, 1650 Orleans Street, Baltimore, MD 21231, USA
HAYWAGA@jhmi.edu

Abstract The genomes of several human herpesviruses, including Kaposi sarcoma (KS) herpesvirus (KSHV), display surprisingly high levels of both genetic diversity and clustered subtyping at certain loci. We have been interested in understanding this phenomenon with the hope that it might be a useful diagnostic tool for viral epidemiology, and that it might provide some insights about how these large viral genomes evolve over a relatively short timescale. To do so, we have carried out extensive PCR DNA sequence analysis across the genomes of 200 distinct KSHV samples collected from KS patients around the world. Here we review and summarize current understanding of the origins of KSHV variability, the spread of KSHV and its human hosts out of Africa, the existence of chimeric genomes, and the concept that different segments of the genome have had different evolutionary histories.

1 Introduction

Kaposi sarcoma herpesvirus (KSHV) (or HHV8) is a gamma-2 or rhadinovirus subfamily herpesvirus (Chang et al. 1994; Russo et al. 1996) that is believed to be the etiological agent of Kaposi sarcoma (KS), primary effusion lymphoma (PEL), and some forms of multicentric Castleman disease in humans. Importantly, both classic KS and long-term inapparent latent infection by KSHV are relatively rare in extant human populations within the USA, Europe, and Asia (approx 1% seropositivity) but are significantly more common in the Mediterranean area (5%–15%) and most common in sub-Saharan Africa (50% or more). Presumably because of reduced immunosurveillance, the incidence of KS increases 500-fold in iatrogenic organ transplant patients and up to 20,000 fold in homosexual AIDS patients in the USA and Europe, and KS is now the most common human tumor found in southern Africa (Chokunonga et al. 1999; Dedicoat and Newton 2003).

Our study of genotypic variation among KSHV genomes derived from different human population groups on different continents has proceeded in three stages. First, we generated a simplistic "ethnic migration" model based only on sequencing of the hypervariable ORF-K1 or VIP genes at the LHS end of the genome (Kajumbula et al. 2006; Zong et al. 1999; Zong et al. 2002). Second, we reported the discovery and analysis of additional complexity and chimerism at the extreme RHS of the genome, associated with the concept of multiple alleles of the ORF-K15 or TMP gene (Poole et al. 1999; Zong et al. 2002) (Zong, J.C., Su, I.J., Morris, L., Alagiozolou, L., Sitas, F., Kajumbula, H., Katange-Mbide, E., Boto, W and Hayward, G.S., manuscript in preparation). Third, we have also now carried out an exhaustive analysis of multiple loci across the central constant segment of the genome (Zong, J-C and Hayward, G.S., unpublished data). Each stage and segment gave rather different results but a reasonably complete picture of the overall pattern of variability within and across the whole length of modern human KSHV genomes has now emerged. These data permit deductions about the history and origin of different parts of the KSHV genome and provide for the first time some important insights into how a large DNA virus genome evolves within the relatively short evolutionary time frame of the divergence of both premodern humanoid species and of modern *Homo sapiens*. In particular, we will introduce the concept of both "old" and "new" segments of the genome and of the role of rare interspecies chimerism, as well as how founder and bottleneck effects associated with modern human migration patterns have impacted the genome.

A critical aspect of this story for the KSHV genome is the realization that the situation observed here is dramatically different from that found in most other more typical herpesviruses (e.g., HCMV). In particular, multiple infections and intragenomic recombination and chimerism are both very rare events in KSHV, presumably because of preferential familial transmission and low horizontal transmission rates, which have in general maintained high levels of linkage between multiple loci across the genome. In contrast, similar multisite analysis of variability and cluster patterns across the HCMV genome has revealed much higher levels of genetic scrambling, with a nearly total loss of subtype linkage between nonadjacent variable loci, apparently resulting from rampant intratypic recombination and extensive penetrance of subtypes across ethnic and geographic boundaries (Zong, J-C et al., unpublished data).

2
KSHV Genetic Variability

2.1
Early Studies of VIP Hypervariability Patterns

Two very different types of variability occur within the VIP (K1) and TMP (K15) proteins encoded at the extreme LHS and RHS, respectively, of the KSHV genome. VIP is a 289-amino acid lytic cycle membrane-associated and ITAM-containing IgG family glycoprotein that displays up to 30% amino acid differences between isolates and clusters into five major subtypes (A, B, C, D, and E) plus numerous minor variants that correlate well with ethnic and geographic ancestry in different human host populations (Biggar et al. 2000; Cook et al. 1999; Fouchard et al. 2000; Hayward 1999; Kadyrova et al. 2003; Lacoste et al. 2000a, 2000b; Lagunoff and Ganem 1997; Meng et al. 2001, 1999; Nicholas et al. 1998; Treurnicht et al. 2002; Whitby et al. 2004; Zhang et al. 2001; Zong et al. 1999, 2002). VIP is unique to KSHV and functions as a Tyr-kinase signaling protein that has antiapoptotic properties, can either mimic or block B-cell receptor activation, and appears likely to play an important role in initial establishment or control of latency (Damania and Jung 2001; Lee et al. 2003, 2005; Wang et al. 2004).

Variability within the VIP gene displays very high nonsynonymous rates, with up to 85% of nucleotide changes creating amino acid changes, implying that some as yet unknown but powerful biological selection process has been involved. Furthermore, this variability can be divided into three distinct levels. First, the oldest level occurs predominantly between subtypes and is distributed relatively evenly across the whole gene. Second, a more recent level is observed even within subtypes but is largely limited to two 40-amino acid blocks referred to as VR1 and VR2. The third level is found even within specific variants but is restricted to a 20-amino acid Cys bridge bounded hypervariable domain (VR*) that resembles the VR3 loop in the HIV ENV protein and appears to correlate with strong T-cell epitopes (Stebbing et al. 2003; Zong et al. 1999, 2002). However, unlike the generation of rapidly evolving pseudospecies of HIV or HCV genomes within a single patient, we have never detected any convincing evidence for multiple variants of KSHV VIP either between multiple KS lesions and PMBC samples or within multiple PEL cell lines derived from a single patient. We have concluded that our evidence for highly ethnically restricted clusters of closely related VIP proteins indicates that (unlike most other herpesviruses) KSHV transmission is predominantly familial, with very low rates of horizontal or multiple infections.

We previously interpreted the patterns of distribution of specific VIP clusters as indicating that the principal VIP subtypes arose during the migration

of modern humans out of east Africa first into sub-Saharan Africa (B branch), then into south Asia, Australia, and the Pacific Rim (D/E branch), and finally into the Middle East, Europe, and north Asia (A/C branch), with very little subsequent remixing (Hayward 1999; Zong et al. 1999, 2002). The divergence of the B branch from the progenitor of all the other branches is judged to have occurred approximately 100,000 years ago, with the split of D/E from A/C occurring close to 70,000 years ago. The current estimated times of divergence of D from E and of A from C are 50,000 and 35,000 years ago, respectively, based on both the overall evidence about human migrations and the lengths of the branches in the VIP phylogenetic trees. Within the A and C branches, individual variant clades (A1 to A10 and C1 to C7) probably arose between 10,000 and 12,000 years ago as human populations expanded out from a limited number of Eurasian Ice Age refuges.

2.2 TMP Allelic Variability and Chimeric Genes

TMP(K15), also known as LAMP, is a 500-amino acid latent-state membrane protein with twelve transmembrane domains encoded by an eight exon spliced mRNA and related to the LMP2 latency protein of EBV. TMP is also a Tyr-kinase signaling protein, although it lacks the ITAM motifs present in both KSHV VIP and EBV LMP2, and by analogy with LMP2 probably contributes to maintenance of the latent state (Brinkmann et al. 2003; Sharp et al. 2002). TMP genes from different KSHV isolates have been described to fall into two alternative allelic subtypes, referred to as P (prototype) and M (minor), that have diverged by 70% at the amino acid level but otherwise show little variability (Choi et al. 2000; Glenn et al. 1999; Hayward 1999; Kakoola et al. 2001; Nicholas et al. 1998; Poole et al. 1999). It is important to appreciate that only a very small segment at the extreme RHS of those genomes that carry the TMP-M allele show this high level of divergence: In fact, it is limited to just the TMP gene itself. However, the patterns of divergence in the region adjacent to the TMP gene (ORF75E and UPS75) are very interesting and complex and provide critical insights and information about the origin and history of the TMP-M allele. Furthermore, about 20% of all human KSHV genomes examined have the TMP-M allele, but they can be found associated with all three of the major branches of VIP genes (A/C, B, or D/E). In essence, there are two overall classes of KSHV genomes, the predominant P class that are considered to have the modern P pattern throughout their length and a second less prevalent class that are chimeric hybrids of P genomes with just a small highly diverged M allele segment at the RHS.

We originally proposed that the TMP-M alleles are of nonhuman origin and are carried only in relatively rare genomes that consist of a predominantly modern human KSHV genome joined to an exotic Old World primate-derived TMP gene at the RHS (Poole et al. 1999). These chimeras were evidently generated by a two-step recombination process, which we suggested involved an initial chimera that was created by a single original cross-species recombination event from an exotic highly diverged primate KSHV-like virus, followed by a second event from a much less diverged premodern humanoid lineage virus, such as might be expected to have been present in Neanderthals, for example. However, there is an alternative possibility that the parent M-type virus from which they originated was instead an anciently diverged true second species of human KSHV virus, which has since either become extinct or has not yet been detected as an intact genome within modern humans (Hayward 1999; Zong et al. 2002).

2.3 More Recent Comprehensive Analyses Including the Central Constant Segment

Subsequent to those original reports, we have continued to extend our detailed analysis of KSHV genome variability in three ways. First, we have sequenced a large number of additional VIP genes (now totaling 180), especially from two major new collections in sub-Saharan Africa, including PBMC from 35 KS patients in Uganda and KS biopsies from 20 patients in South Africa. Second, we have sequenced a total of 60 complete TMP genes from a selected subset of the same samples evaluated previously for VIP, as well as from many of the new African samples. Third, we have also carried out PCR sequencing at up to 10 additional internal genomic loci in many of these same genomes (averaging between 60 and 150 samples each).

Interpretation of the results of these new data from all three segments of the genome now permits refinement and expansion of our previous models, leading especially to the concepts of both chimeric herpesvirus evolution and differences in the recent history of the "new" LHS and "old" RHS ends of modern human KSHV genomes. Based on the overall picture that has emerged, we can now classify known KSHV genomes into a model involving "12 principal genotypes" (see Fig. 1), in which we view each genotype class as being composed of variable-length segments with different estimated ages of divergence that correspond to the different degrees of variability or sequence conservation found at the 12 PCR loci studied. Unlike the hypervariable VIP and TMP genes, the level of variation at the 10 internal constant region loci is much smaller (on the order of 1% to 5% nucleotide differences only), and in this case the polymorphisms rarely affect the primary protein structure,

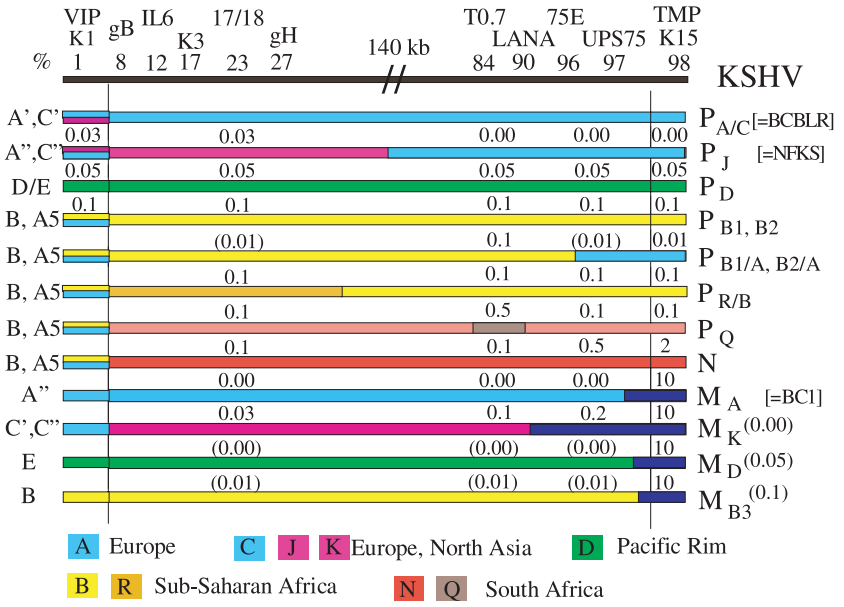


Fig. 1 The 12 major KSHV genotypes model. The diagram illustrates the deduced segmental and chimeric patterns of the 12 principal KSHV genotype structures. The linear genomic positions (%) of the PCR loci used are depicted at the top above the genome lines. For simplicity, the central segments (including the RAP/K8.1 locus at position 37) have been omitted (//). The allelic and subtype descriptions are given to the *right*, and the nine segment subtype color codes are denoted *below* the diagram. In addition, the exotic RHS M allele segments are designated by the *dark purple bars*. The estimated ages of divergence from the prototype Eurasian P genome on the *top line* are given for each major segment of each genotype in millions of years ago (MYD values)

but nevertheless they still show parallels to the subtype clustering patterns observed in VIP and TMP. Where possible, we have tried to obtain data for the complete VIP gene, not just fragments of it, and to analyze nearly all of the internal loci for each genome to obtain a complete and representative sampling picture across each genome.

3 Phylogenetics of the Hypervariable VIP or K1 Genes

A stylized generic radial phylogenetic tree showing the major branching and cluster patterns and divergence ages of VIP proteins from major human

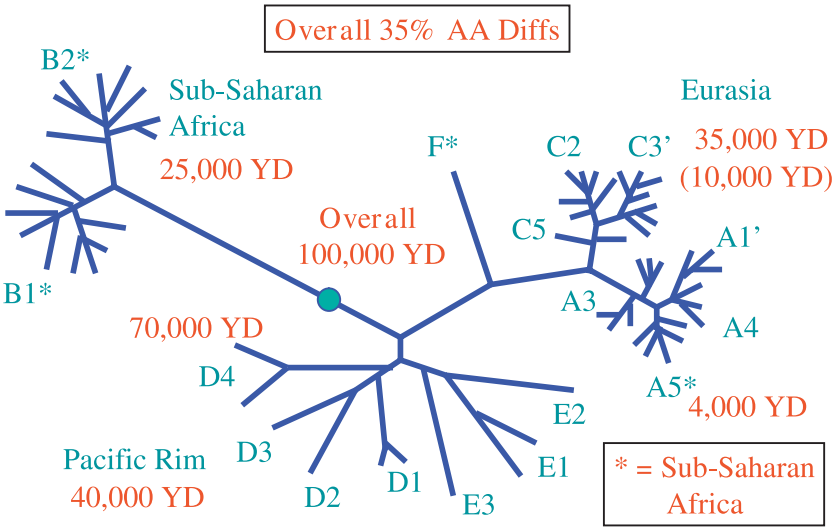
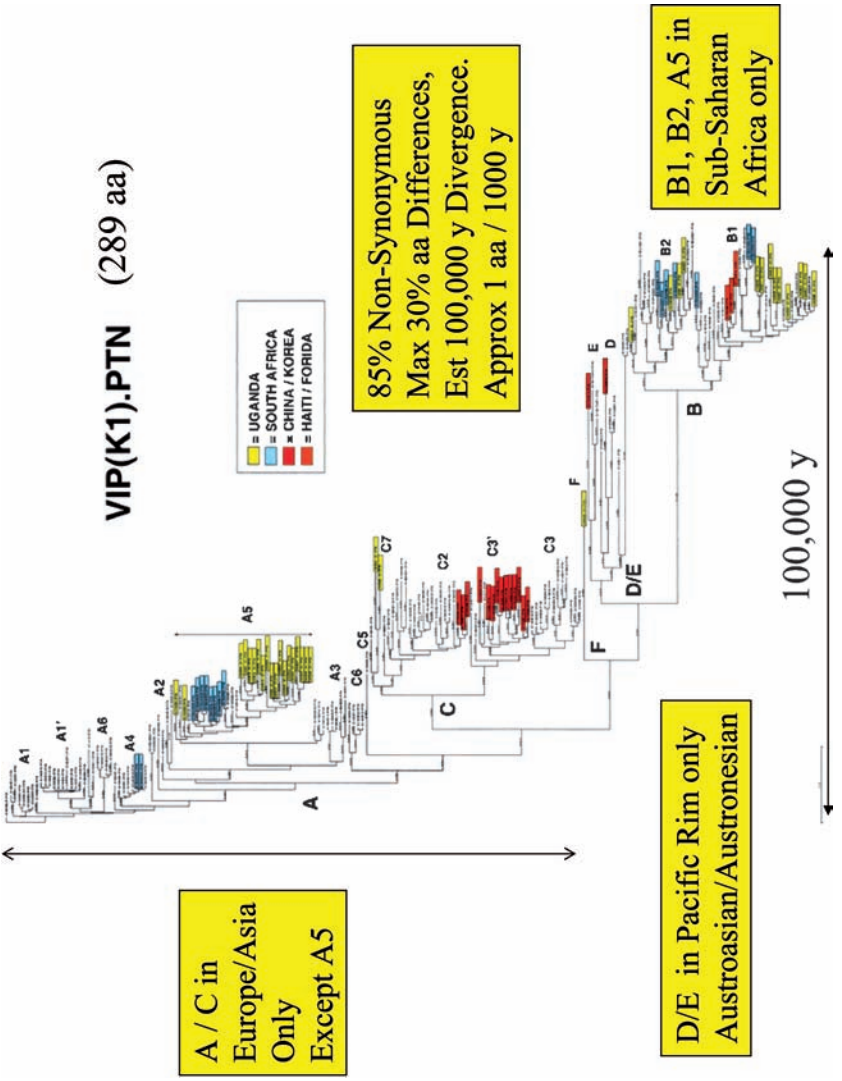


Fig. 2 Overall radial tree branching pattern for the KSHV VIP(K1) hypervariable ITAM membrane signaling protein. The diagram summarizes the major features of clustering and relative distance and branching patterns for samples from Eurasia, the Pacific Rim, and sub-Saharan Africa. The *asterisks* denote the four VIP subtypes found only in sub-Saharan Africa (B1, B2, A5, and F)

Fig. 3 Detailed linear phylogenetic tree for all 180 VIP proteins analyzed in our studies. This tree is based on the complete 289-amino acid VIP/K1 ORF. The entire tree is judged to contain only "modern" VIP subtypes, representing no more than 100,000 years of divergence. Overall, the VIP protein displays 30% amino acid variability, with 85% of the nucleotide polymorphisms being nonsynonymous. This gives an estimated rate of change of approximately one amino acid per 1,000 years. *Color patterns* highlight the samples from Uganda (*yellow*), South Africa (*blue*), Chinese Taiwan and Korea (*pink*), and American samples with African connections (*green*). The two Taiwan samples in the D and E branches were from aboriginal Hwalian samples. HKS22 from Uganda represents the sole example of an F subtype VIP gene found. In general, the A and C subtypes are found only in European and Asian samples, whereas B subtypes are found only in sub-Saharan African populations, and D and E subtypes are found only in Austronesian/Austroasian branch populations including Polynesians and Amerindians

population groups is presented in Fig. 2. The actual linear phylogenetic VIP protein tree for all 180 KSHV samples that we have sequenced plus a few representative samples from other studies is shown in Fig. 3, with key ethnic subgroup collections indicated in the color key.


VIP(K1).PTN (289 aa)
= UGANDA
= SOUTH AFRICA
= CHINA / KOREA
= HAITI / FORIDA
85% Non-Synonymous
Max 30% aa Differences,
Est 100,000 y Divergence.
Approx 1 aa / 1000 y
B1, B2, A5 in
Sub-Saharan
Africa only
A / C in
Europe/Asia
Only
Except A5
D/E in Pacific Rim only
Austroasian/Austronesian
100,000 y
A1
A1'
A6
A4
A2
A5
A3
C6
C5
C7
C2
C3'
C3
A
C
F
D/E
E
D
B
B2
B1

3.1
Age and Origin of B-Subtype VIP Genes in Sub-Saharan Africa

In our original analysis of 60 complete VIP genes from around the world (Zong et al.), eight of nine samples from sub-Saharan African were B subtypes, including our original B prototype (431K), and one was the prototype A5 variant (OKS3). Only five other examples of B-subtype VIP genes were found at that time, but four of these were also from patients of African origin, including ZKS6 and JKS20 from African emigrants to New Zealand and the United States, respectively, JKS15 from an African American patient in Baltimore, and OKS7 from a Haitian patient in Miami, Florida. OKS8, which closely resembled OKS7, came from a Mexican woman with AIDS in Miami, and it remains the only exception that we have found of a B-subtype VIP gene without an obvious direct connection to Africa. A total of 30 additional complete B-subtype VIP genes have now been sequenced and the overall results are shown in the partial VIP phylogenetic tree given in Fig. 4. All but one of the B-subtype genomes came from sub-Saharan Africa (either Uganda or South Africa), with the exception being the VG1 PEL sample from Haiti. Data for an incomplete internal segment of the VIP genes from the same 23 HKS samples described here were also presented recently in an analysis of just the VIP genes from patients with different tribal affiliations from Uganda (Kajumbula et al. 2006).

Overall, the 39 B-subtype VIP proteins that we have analyzed, although all differing from the A/C Eurasian subtype by 30%, show only moderate intratypic variations (totaling 8% at the amino acid level), and they cluster into two very distinctive major branches that we refer to as B1 or B2 variants. 431K is designated as our prototype for the B1 VIP gene, and SAPB3 is designated as our prototype B2 VIP gene. There were 13 B1s and no B2s among our original set of 14 examples of African VIP genes, which included five of five from Gambia, three of four from Tanzania, two from Uganda, and three from USA or Haiti (Zong et al., 1999). However, we have now found that four of eight Bs from South Africa and five of 15 Bs from Uganda proved to be B2 variants rather than B1 variants. Several of the new African VIP genes display novel in-frame deletions within the VR1 or VR2 regions. From the analysis of the partial Uganda VIP genes, the possibility that HKS9 and HKS27 represent a distinct third B3 subgroup was suggested in our earlier analysis (Kajumbula et al. 2006), and this remains plausible for the complete genes, including the South African samples in our larger analysis.

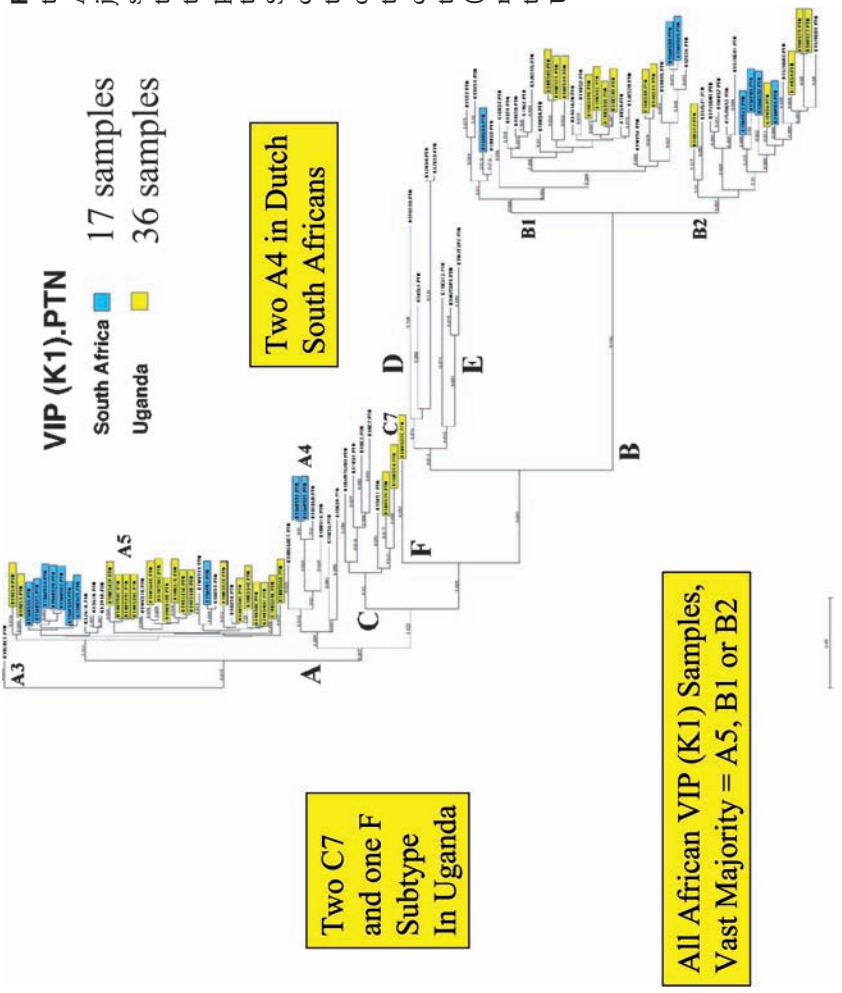


Fig. 4 Detailed linear phylogenetic tree of the Ugandan and South African VIP proteins. The vast majority of all sub-Saharan African samples are B1, B2, and A5 subtypes. The 36 complete VIP proteins from Uganda (PBMC from KS patients) are shown in *yellow*, and the 17 complete VIP proteins from South Africa (KS biopsies) are indicated in *blue*. The diagram illustrates the relatively wide divergence of B1 and B2 branches compared to the very narrow divergence of the A5 clade. The relative branching positions of the five anomalous samples (two A4 from patients of Dutch European descent in South Africa and the two C7 and one F subtype from Uganda) are also indicated

3.2 The African A5 Cluster of VIP Genes

In addition to the characteristic African B-subtype VIP genes, we have also now detected 34 more of the A5 subtype, plus one novel F- and two novel C7-subtype VIP genes from Uganda and South Africa (Fig. 4). In addition, two A4 examples from South Africa represent the only classic European variants that we have found in Africa, but these are the exceptions that prove the rule, because both came from the only known Caucasian patients among the group, who were both South African men of Dutch descent. In contrast, the rare prototype examples of VIP-F (HKS22) and the C7 subtypes (HKS35, HKS54) from Uganda all appear to be of genuine endemic African origin. Surprisingly, A5 variants have been detected relatively uniformly throughout Africa, including within a majority of our samples from South Africa as well as many from Uganda.

The 35 A5 VIP genes that we have analyzed all fall into a very tight and narrowly diverged cluster that is quite distinct from the closest European-north African variant clusters A3 and A7. However, we have never found an A5 subtype VIP gene outside of sub-Saharan Africa. Overall, the clustering of A5-subtype VIP proteins is extremely high, with branch lengths on the phylogenetic tree that are very short, being less than 1/5th that of both the B1s and B2s, suggesting a common origin no more than 4,000 years ago, with the B1 versus B2 (Fig. 4) divergence probably occurring about 25,000 years ago.

One other early study reported only A5-subtype VIP genes in PBMC among 15 children with symptomatic fevers in Zambia (Kasolo et al. 1997), but in retrospect it seems extremely unlikely to us that the particular primers used in that study would have been able to detect B-subtype VIP genes. Therefore, the implication that A5s may predominate in currently circulating viruses associated with lymphadenopathy and febrile illnesses in children, rather than in adult KS (although plausible), needs to be re-evaluated. The A5 VIP proteins are clearly much more similar to Eurasian VIP genes than to the B-subtype VIP genes, and they must have had a single-source introduction from Europe or the Middle East presumably via North Africa (backwards flow). The A5s evidently have been (and presumably continue to be) spreading rapidly across and throughout sub-Saharan Africa over a time frame that appears to coincide with the Bantu expansion from western Africa over the past 4,000 years. It is also important to emphasize that all KSHV genome samples that we have studied that have A5-subtype VIP genes are recombinant chimeras in which only a very small LHS segment is A-like, and the remainder of their genomes consist of typical distinctive sub-Saharan B, Q, R, or N subtypes (Zong et al. 2002) (Zong, J-C. et al, manuscript in preparation).

3.3
Source and Origin of the Novel Ugandan VIP-F and VIP-C7 Genes

Among the total of 69 complete VIP genes that we have studied with clear sub-Saharan African origins, only three, namely HKS35 (C7), HKS54 (C7), and HKS22 (F), which are all from Uganda, do not fit into the B1, B2, or A5 patterns described above. HKS35 and HKS54 cluster together between the two major groups of subtype-C VIP proteins, namely, C″(C3) and C′(C1/C2), and are referred to as the first examples known of C7 variants. These are closest to the C4 and C5 variants from Saudi Arabia (with an estimated divergence from other Cs approximately 15,000 to 20,000 years ago). In contrast, HKS22 lies between the A/C and D/E branches on the VIP protein phylogenetic trees (Figs. 2–4) and is the only sample that we have ever detected that falls outside of the three major branches. It is therefore tentatively referred to as an F-subtype VIP gene, suggesting divergence from the other main branches approximately 50,000 years ago. However, the C terminus of HKS22 is very similar to the C7 pattern, including the diagnostic five-amino acid deletion common to all Cs, and this sample clearly represents a complex F/C intragenic recombinant. In addition to HKS22, there are several other samples known (all from central Africa) that also do not fit readily into any of the currently defined A to E classification patterns. These include SAN2 (Whitby et al. 2004) and UGD23 (Kakoola et al. 2001), plus 43/Ber and 8/Dem (Lacoste et al. 2000b). Although they do differ very significantly from one another, we suggest that these, together with HKS22, are all members of a new fourth main branch that also splits into two subgroups designated F and G.

Finally, there is the question of the source or origin of these rare C7, F, and G VIP variants. In fact, all three genomes containing C7- or F-subtype VIP genes are extremely unusual at all loci tested, being regarded as novel intermediate F or G subtypes at several internal constant region loci (see Fig. 8; Zong, J-C et al., unpublished data) and either having unusual divergent B4- and B5-subtype ORF75 and TMP genes (HKS54 and HKS22) or, in the case of HKS35, having N-subtype ORF75 and TMP genes. We presume that the VIP C7 genes were relatively recently reintroduced into sub-Saharan Africa from the Middle East or Europe, whereas the novel F (and G) subtypes probably represent one or more genuine independent branches that may either never have left Africa or more likely returned into Africa from an early Middle Eastern branch that diverged after the south Asia/Pacific Rim branch but before the Eurasian A/C branch.

3.4
A Novel E-Subtype VIP Gene from a Taiwan Aborigine

Several rare KSHV VIP protein sequences have been described from indigenous populations around the Pacific Rim. Although diverged from one another by up to 24%, these nevertheless cluster into either of two distinctive D and E branches within a single D/E superbranch that is intermediate between the Eurasian A/C and African B branches (Figs. 1–3). Examples of D-subtype VIP genes include one (D1) from a Taiwan Hwalian aborigine (Zong et al. 1999), two (D2) from Pacific Island Polynesians seen in New Zealand (Zong et al. 1999), one (D3) from Australia that is presumed to be of aboriginal origin (Meng et al. 1999), and three (D4) of presumed Ainu origin from Hokaido in Japan (Meng et al. 2001). In addition, two VIP genes from Tupi Amerindians in Brazil were designated as the first examples of a new VIP subtype E1 (Biggar et al. 2000). More recently, Whitby et al. (Whitby et al. 2004) have also described five more partial VIP sequences (E2) from two Ecuadorean Amerindian tribes that branch close to the Tupi samples.

Interestingly, a second Hwalian sample (TKS13) that we have analyzed also clustered with the nearly identical D1 (TKS10)- and D2 (ZKS3)-subtype genomes in both the ORF26 and T0.7 constant region loci (Zong, J-C. et al., manuscript in preparation) but proved to have a VIP protein that is more similar to the E subtypes than to TKS10 or any other Ds. Therefore, we propose to designate TKS13 as E3 in contrast to the Brazil (E1) and Ecuador (E2) Amerindian subtypes. Furthermore, on the RHS, TKS13 proved to have a TMP-M allele, whereas TKS10 and the two Polynesian samples (ZKS3, ZKS4) have TMP-P alleles (see Fig. 5 and below). Unfortunately, sequence data for the constant region loci and TMP are not available for comparison from any of the other known Pacific Rim KS genomes with D or E VIP genes. Overall, we interpret that these 14 genomes with D/E pattern VIP genes were all derived from human populations that have common ancestors who initially migrated out of Africa approximately 70,000 years ago as a single distinctive south Asian branch, giving rise to all subsequent native Australian, Austroasian, Austronesian, and Polynesian populations, as well as Ainu and Amerindian branches, that is, collectively representing a Pacific Rim branch of modern human KSHV. Furthermore, all seven individual variant lineages within the D and E subtypes exhibit much longer branches than any of the variant clusters within the A, C, or B branches (Figs. 1–3), and therefore we suggest that all of these diverged from one another at least 40,000 years ago.

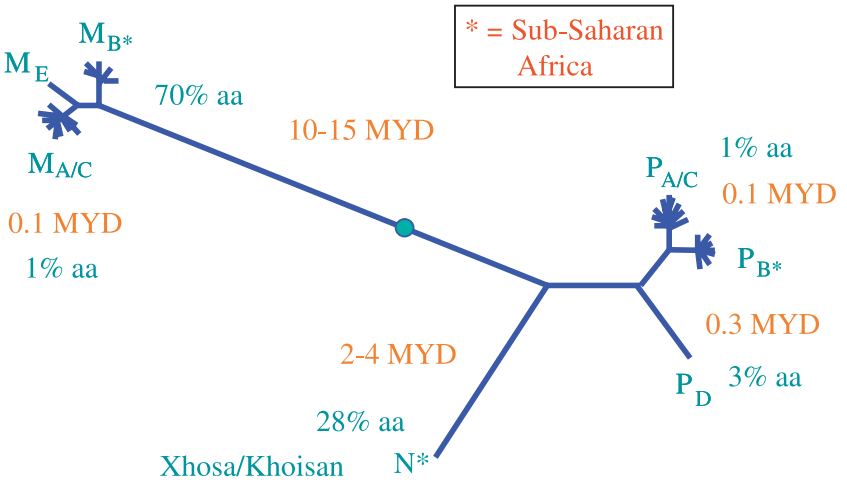


Fig. 5 Overall radial tree branching pattern for the KSHV TMP(K15) membrane signaling protein. The diagram summarizes the clustering, relative distance, and branching patterns for samples from Eurasia, the Pacific Rim, and sub-Saharan Africa. Estimated divergence times in millions of years (MYD) are given in *red*, and the overall % amino acid difference levels are shown in *blue*

4 Variable Sequence Loci within the Constant Region

To contrast with the hypervariable LHS and RHS termini of the KSHV genome, we refer to the remainder of the genome as the constant region. However, there are patches of low-level nucleotide variability all the way across the genome, and we have selected known sites of variability within the constant region as likely to be the most informative. The specific internal loci chosen for PCR sequencing analysis range in size from 640 bp up to 1,900 bp and include similar or extended versions of four previously utilized loci (ORF26E, T0.7, ORF75E, and UPS75) as well as five new loci encompassing segments of the vIL6, vMIR2/K3, ORF18/19, RAP/K8.1, and LANA1 genes. Some data have also been obtained for the UL22(gB) locus, but other potential loci such as the ORF4, vIRF1, vCYC, and vGPCR genes showed much less variability still. All of these nine chosen PCR loci together represent a sampling of 7,823 nucleotides (or 5.6% of the 137,000-bp unique segment genomes). Across this sampled part of the constant region we have detected a total of 409 chimeric nucleotide positions (5.2%), with the differences between a prototype Eurasian P(A/C) genome and the prototype P(B1) genome totaling 86 nucleotides (1.1%).

4.1
Clustering, Subtyping, and Linkage

We then interpreted these results in terms of clustering and subtyping at each locus, as well as in terms of maximal levels of linkage between loci and minimal inferred chimerism. Based on the concept of prototypic nonchimeric genomes assigned for each major subtype [e.g., BCBL-R as a average typical P(A/C) genotype, 431K as an average typical P(B)-subtype genome, and ASM72 as the most typical prototype P/M chimera], we have been able to define matching linked subtype patterns at all loci with considerable confidence. Major clustered deviations from these prototype subtype patterns at each locus were then assigned additional subtype status and matched between loci, using the assumption of maximal linkage. Deviations from the standard average or commonest linkage patterns are then considered evidence for additional chimeric recombination events.

A major feature of the results at all variable gene loci for KSHV (as well as for HCMV) is that the sequences show a very high degree of clustering, whereby almost all samples fall into one or another of a relatively small number of subtypes at each locus that may differ from each other quite extensively but themselves show very low levels of variability. There are occasional examples of intragenic chimerism with very obvious recombination junctions, but it is quite clear that there is not a continuum of variability or an infinite number of subtypes. However, the level of variability between subtypes can vary significantly between the different loci and the number of distinctive subtype branches found can also vary (especially in HCMV).

4.2
Subtype Distribution Patterns Within the Constant Region

One of the most puzzling aspects of the original VIP gene analysis (Zong et al. 1999), was that although it very neatly matched the expected patterns of strain divergence corresponding to founder effects from the migrations out of Africa, it did not display the anticipated patterns of greater diversity within sub-Saharan Africa than elsewhere that would have been expected based on the overall mitochondrial and Y chromosome phylogenetic trees of modern humans. However, this was not the case for the constant region loci, where much greater and older divergence is indeed found within the samples from sub-Saharan Africa. For example, we now recognize and define nine distinct subtype clusters within the T0.7 locus. With a few key exceptions, three of these are almost exclusively found only within Eurasian KSHV genomes (A/C, J, and K/M), whereas the D/Es are found only in aboriginal Pacific Rim samples and the other five (B, F/G, R, Q, and N) are found exclusively within sub-Saharan

African samples. All other constant region loci are also much more similar to this cluster distribution pattern than they are to the VIP patterns.

A summary of the observed levels of nucleotide divergence between subtypes at each of nine different constant region loci is presented in Table 1. Note that this analysis omits the hypervariable VIP and TMP loci and therefore greatly underestimates the overall level of variability between subtypes. There are three initial points to be made. First, within the constant region sites (especially at the RHS) there is essentially no distinction at all between genomes that have the A- and C-subtype VIP genes, and therefore we refer to them all as A/C. Second, there are nevertheless two other distinct subtypes within Eurasian constant region loci, which we refer to as J and K. Both are intermediately diverged compared to A/C and B, but there is a pattern whereby the "older" J and K constant region subtypes are preferentially linked to the VIP superpatterns referred to as A′ and C′, whereas the "newer" A/C subtype constant regions are preferentially linked to the A″ and C″ VIP superpattern genes (Zong et al. 2002). Furthermore, essentially only the K subtypes show direct linkage to the RHS TMP-M alleles. Third, it is evident that although the F- and G-subtype constant region genes are generally less diverged from A/C than the B patterns, the novel R, Q, and N subtypes are most often even further diverged from the Eurasian subtype than the typical African B pattern genomes.

4.3 The 12 Major Genotype Patterns in KSHV

Combining the results of our analyses for all three regions (i.e., the LHS VIP gene, the central segment constant region loci, and the RHS TPM alleles), leads to an overall picture that can be summarized in a highly simplified format by dividing the genotypes into 12 very distinctive and representative patterns as shown in Fig. 1. The subtype designations of different domain lengths of each genome are illustrated by the colored bars, with dark purple bars denoting the exotic TMP-M alleles. We envisage each of the different genotypes here as being composed of multiple segments with different ages of divergence. Numerical estimated ages (in millions of years) for each segmental domain are given in terms of the timing of its original divergence from the predominant European A prototype pattern, which is presented on the top line. Overall, there is a dramatic trend from left to right across the genome of finding relatively young segments only at the extreme LHS, toward retention of medium-age to older segments as well in the center, and then with some much older segments included as well toward the extreme RHS of the genome. The major features that contribute to this model of the 12 prototype genomes are described in the sections below.

Table 1 Comparison of the number of constant region nucleotide polymorphisms detected among KSHV genotypes

						Eurasia			Pacific Rim	Sub-Saharan Africa									
Locus name	Map position	Length (bp)	Sample No.	Polymorphisms		A/C vs. CM J-A/C	A/C cdvs. K-M	J-A/C vs. K-M	A/C vs. D	A/C vs. F	A/C vs. G	A/C vs. B	A/C vs. R	A/C vs. Q	A/C vs. N	B vs. N	B vs. K-M	K-M vs. N	B vs. G
				Total	Eurasia														
vIL6	13.3%	1826	69	75	19	11	13	2	5	19	21	22	20	17	20	19	26	23	21
K3	14.4%	920	71	48	12	3	3	2	12	NA	NA	10	9	12	12	5	6	9	NA
18/19	23.5%	930	51	28	8	8	8	0	11	NA	NA	13	13	8	13	10	11	7	NA
26E	34%	950	73	36	15	5	9	8	4	3	NA	11	8	9	11	12	15	13	NA
K8/8.1	58%	1900	48	62	31	NA	29	NA	13	NA	32	35	23	23	28	15	13	8	12
T0.7	84%	597	149	65	19	3	10	8	11	7	14	11	12	24	18	11	11	12	10
LANA**	90%	980	95	61	15	2	9	9	6	1	6	5	6	3	15	12	11	20	8
75E	96%	850	116	34	16	0	12	12	6	7	9	7	9	12	13	10	10	2	6
UPS75*	96.5%	770	116	66	18	0	16	16	17	7	6	7	NA	8	44	37	26	52	3
Total		9723	161	475	153	32	109	57	85	44	88	121	90	116	175	131	129	146	60

*Truncated before the hypervariable TMP(K15) C terminus; **nonrepetitive N terminus; NA, not known yet

4.4 Geographic Distribution of Constant Segment Subtypes

It is also dramatically evident that, just as for VIP, there are major differences in the distribution of constant region subtypes between the Eurasian samples (Fig. 6) and those from the Pacific Rim (Fig. 7, upper group) and sub-Saharan Africa (Figs. 7 and 8). In fact, in most cases there are also major differences

KSHV – EURASIAN GENOTYPES

		K1	gB	vIL6	K3	18/19	26E	T0.7	LNA	75E	UPS	K15	
BCBL-R		A1	A/C	A/C	A/C	A/C	A/C	A/C	A/C	A/C	A/C	P	
BCP1	*	A1	A/C	A/C	A/C	A/C	A/C	A/C		A/C	A/C	P	
BCBL-B		A4	A/C	A/C	A/C	A/C	A/C	A/C		A/C	A/C	P	
2x SAKS21		A4	A/C	A/C	A/C	A/C	A/C	A/C	A/C	A/C	A/C	P	HOLLAND
EKS1	*	o C2					A/C	A2		A/C	A/C	P	EGYPT
BKS12	*	o C2					A/C	A2	A2	M	M	MX	
BC1		A2	A/C	A/C	A/C	A/C	A/C	A2	A2	M	M	MX	
BKS16		A6						A6	A2	M	M	MX	
BC2		C3	A/C	A/C	A/C		J	A/C	A/C	A/C	A/C	P	
JSC1		C3	A/C	A/C	A/C		J	A/C		A/C		P	
BKS15		C3					J	A/C		A/C		P	
2x RGKS3/4	*	C3			K		A/C	A/C		A/C	A/C	P	SICILY
2x BKS3/4	*	C3					J	A/C				P	
ASM-K3		A1		J	A/C		A/C	A/C	A/C	A/C	A/C	P	
OKS13		C3					A/C	A3	M	--	M	MX	
SKS3	*	o C2					K	A3	M	M	M	MX	SAUDI
SKS9	*	o C5					K			M		MX	SAUDI
BC3		C3	A/C	J	K	J/K	J	A/C		A/C		P	
ARK1	*	o A7	J/K	J	J		A/C	A/C	A/C		A/C	P	ITALY
BCBL1	*	o A3	J/K	J	J	A/C	A3	A3	A3	A3	A/C	P	
OKS18	*	o C2						C2			A/C	P	
KSF		C3	J/K	J	K	J/K	A/C	A/C	A/C				GERMANY
BKS20	*	o A7					K	A/C			A/C	P	
11x KKS5	*	C3'			J	J/K	K	M	M	M	M	MX	CHINA/K
3x KKS6	*	o C2	J/K		K	J/K	K	M	M	M	M	MX	CHINA/K
BKS21	*	o C2				J/K		M		M	M	MX	
BKS13		o C1					K	M	M	M	M	MX	
ASM70		o C1	J/K	K	K	J/K	K	M	M	M	M	MX	

* = Classical, non HIV associated o = Inner, older A' + C'

Fig. 6 Whole genomic structural summary for 43 Eurasian KSHV samples, including 23 P/M chimeras. Color coding distinguishes the M-associated segments (*red*) from P(A/C) patterns on the RHS, as well as the intermediate-aged P(J) and P(K) subtype patterns (*pink*) from the P(A/C) patterns on the LHS. *MX* indicates the exotic M TMP allele. The chart diagram is based on direct PCR data for up to 10 loci across each genome. Multiple related samples that have the same or very similar patterns are shown together on a single line

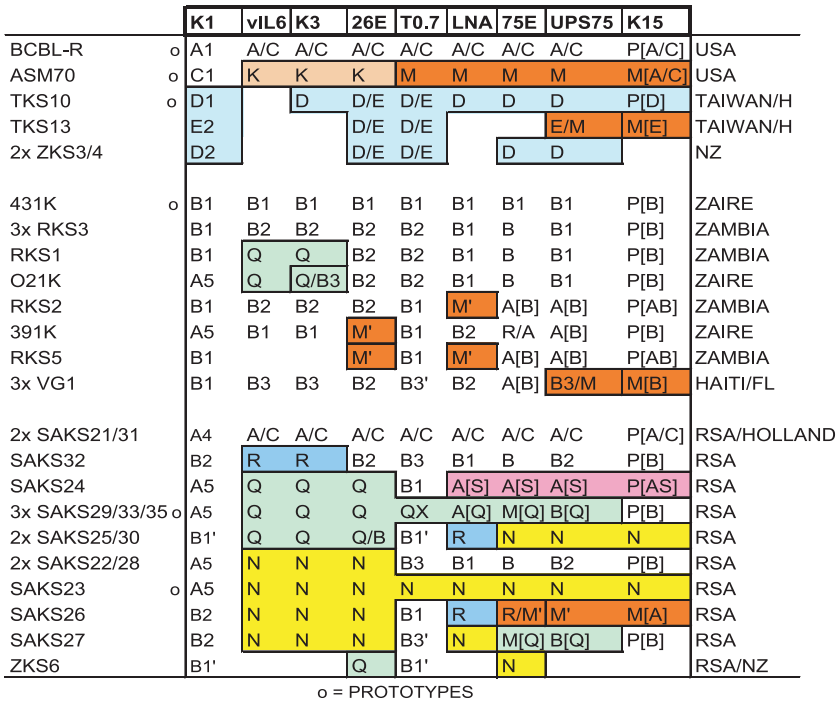


Fig. 7 Whole genomic structural summary comparing four Pacific Rim KSHV samples (*upper group*) with the 15 South African samples (*lower group*) plus 14 other non-Ugandan samples (*central group*). Color coding distinguishes novel D, E, K, M, M′, and Q segments from prototype A/C or B patterns in the *top* and *center* groups, respectively, and distinguishes the A[S], A[Q], Q, R, M, and N segments from prototype A/C or B patterns in the *lower group*. *RSA*, Republic of South Africa. The three prototype PA/C(BCBL-R), PB(431K), and P/M(ASM70) genotypes are also included

between the subtype patterns found in our South African samples (Fig. 7, lower group) and those found in Uganda (Fig. 8) or other places in central Africa (Fig. 7, center group). In particular, both LHS and RHS N-subtype segments were found almost exclusively in South Africa, and LHS R-subtype segments were found primarily in Uganda, whereas LHS Q patterns were common throughout Africa.

One of the most notable features of sub-Saharan KSHV genomes compared to Eurasian and Pacific Rim genomes is an evidently much greater degree of chimerism within the central constant segment region, with many genomes that fall into just one of the “African” classes, in fact differing extensively

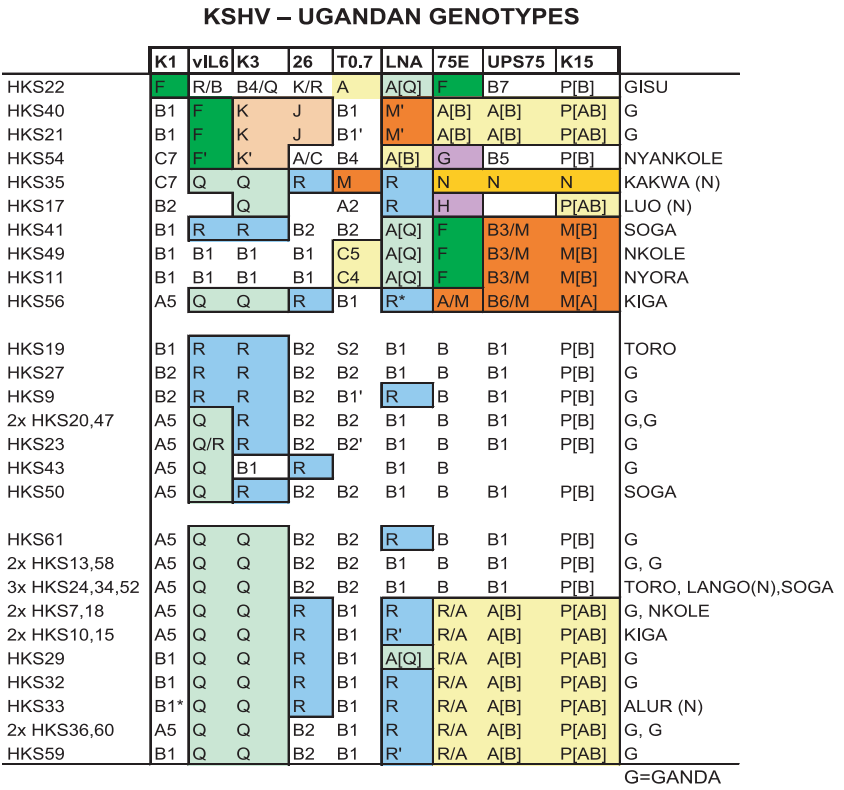


Fig. 8 Whole genomic structural summary for 34 Ugandan KSHV samples. Color coding distinguishes the segments designated as subtype F, G, H, J, K, Q, R, M′, and N from the commonest prototype sub-Saharan African B subtype pattern. Eurasian-like segments designated A, A[B], A[Q], and C are shown in *yellow*. M-allele and M-associated RHS segments are given in *red*. Other more complex chimeric segments are also indicated (see HKS22 especially as the most extreme example). Tribal affiliations are also listed on the RHS (Kajumbula et al. 2006). These genotypes are clustered into three distinct categories in the *upper, central, and lower groupings*, with the latter two being commonest within the Ganda tribe (*G*)

in the size, patterns and complexity of internal chimeric segments that mix Q, N, F, G, R, or even M′ loci into otherwise B genome backgrounds, for example. This is particularly common within a subset of Ugandan genomes derived from non-Ganda tribal sources (Fig 8, upper group). In contrast, the genomes from the Ganda tribe, the predominant Bantu group in Uganda, were of two major types, either those with mostly the common B-subtype patterns throughout except for R or Q at the LHS (Fig. 8, center group) or those with a more complex Q, B, R/A[B] chimeric pattern (Fig. 8, lower group).

5
The TMP-P and TMP-M Alleles and Adjacent Chimeric Regions

Variability in the KSHV TMP(K15) membrane protein mapping at the extreme RHS of the genome represents an entirely different situation than that for VIP. There are huge amino acid variations found between the P and M alleles, but these are not directly associated with the ethnic and geographic diasporas of modern humans. Rather, each allele itself displays the migration out of Africa patterns at a low variability level, and the differences between the P and M alleles (plus now a third newly recognized N allele as well) represent instead a vastly older evolutionary divergence in a gene that is not itself hypervariable like VIP but evidently evolves at the same slower rate as a typical constant region gene. Our reasons for reaching this conclusion are outlined below, and the current picture of TMP branching and divergence levels is summarized in Fig. 6 and Table 2.

5.1
Relative Lack of Variation Within UPS75 and the TMP Coding Regions of the Eurasian Subtypes of the TMP-P and TMP-M Alleles

Although the 15% differences between A and C VIP subtypes probably arose when modern humans first migrated as two major branches into Europe and

Table 2 Nucleotide differences and estimated divergence rates in TMP alleles

A. Other TMP-M Variants compared to the protype TMP-M strain (HBL6/BC1)

Name	Total Nucl diffs (2,100)	Exon Nucl diffs (1,500)	Intron Nucl diffs (605)	Total Amino acid diffs (500)	Est. years of divergence
VG1(B1)	27 (1.3%)	16 (1.0%)	10 (1.6%)	10 (2.0%)	100,000
HKS49(B2)	29 (1.4%)	18 (1.1%)	11 (1.8%)	12 (2.4%)	110,000
TKS13(E)	17 (0.85%)	8 (0.55%)	9 (1.5%)	6 (1.2%)	60,000

B. Other TMP-P variants and TMP-N compared to the prototype TMP-P strain (BCBL-R)

Name	Total Nucl diffs (2,080)	Exon Nucl diffs (1,470)	Intron Nucl diffs (600)	Total Amino acid diffs (489)	Est. years of divergence
431 K(B)	19 (0.95%)	13 (0.9%)	6 (1.0%)+	7 (1.4%)	100,000
TKS10(D)	47 (2.3%)	33 (3.3%)	14 (2.3%)+	25 (5.1%)	250,000
SAKS23(N)	425 (21.3%)	287 (19%)	138 (23%)*	140 (28%)	2,000,000

* Plus severalnon-contiguous blocks; + 10-bp insert

northwest Asia about 35,000 years ago, the 6% to 9% amino acid variations between and among individual variants of the subtype A and subtype C VIP branch has clearly arisen much more recently, probably within the past 10,000 to 12,000 years since expansion out of Ice Age refuges. In contrast, the rate of change within the TMP gene over this same time period has evidently been very slow and certainly is no greater than that in other typical constant region genes (Zong et al. 2002) (Zong, J.C. and Hayward, G.S., unpublished data). For example, there are very few differences at all (a total of only 8 nucleotide changes over 2500 bp) among nine representative TMP-P genes that we have sequenced from both European and Middle Eastern sources that are associated with a variety of A- or C-subtype VIP genes (Fig. 6). This group also includes the European GK18 sample described by Glenn et al. (Glenn et al. 1999). Similarly, across an 850-bp segment of the adjacent ORF75 gene (UPS75 locus) from a total of 17 genomes from Eurasian sources with TMP-P alleles, there were again no more than three total nucleotide changes detected, irrespective of what variant of either A- or C- subtype VIP genes were present (all those shown as P in Fig. 6). Similarly, among African UPS75 genes there was a surprisingly large subgroup of 22 examples with an A/C-like pattern, referred to as the A[B] subgroup (Figs. 7 and 8). Most of these display just a single distinctive additional nucleotide change from the Eurasian A/C pattern here, although at least nine have a common chimeric junction linking them to the African R subtype at the adjacent ORF75E locus.

In contrast, over 30% of all Eurasian KSHV samples examined, including 15 of 16 from Taiwan Chinese and South Korean KS patients, have the hugely different M-allele version of TMP (labeled M[X] and shown in red in Fig. 6), which is 70% diverged from the P-allele at the amino acid level (Poole et al. 1999; Zong et al. 2002). However, once again, sequencing of the complete coding regions plus adjacent 5′- and 3′-flanking areas from five TMP-M genes from both European and Asian sources showed no nucleotide differences at all over nearly 2,500 bp relative to the prototype BC1/HBL6 TMP-M gene. Similarly, within the nearby conserved UPS75 K/M locus (between positions 840 and 1300), 13 Eurasian genomes with associated TMP-M alleles displayed 16 common nucleotide differences (3%) from the TMP-P-associated UPS75 (A/C) prototype but again contained a total of only one variable position among them. Two other subtype K/M UPS75 genes derived from South Africa (SAKS26, FTKS2) were also identical to one another here but showed three differences from the Eurasian versions (Fig. 8 and see below).

5.2 Distinctive African Versions of TMP-P

Despite the nearly total lack of variation within and between A and C subtypes of both the Eurasian TMP-P and TMP-M alleles, there are consistent patterns of nucleotide differences within both the TMP-P and TMP-M allele versions that originate from sub-Saharan Africa and from south Asia/Pacific Rim sources. Our prototype African versions of TMP-P (431K) and TMP-M (VG1) were sequenced across the complete TMP coding and flanking region (2,500 bp) and proved to display a total of 18 and 29 nucleotide changes, respectively (representing 0.8% and 1.2%), compared to the prototype Eurasian TMP-P (C282) and TMP-M (BC1/HBL6) allele genes. Overall, with the exception of the few novel N-allele forms (see below), all 40 sub-Saharan African TMP-P allele genes examined proved to be either typical B subtypes similar to 431K (22 examples) or typical A/C-like Eurasian subtypes similar to BCBL-R (18 examples). However, there was considerably more sporadic variation among the African B-subtype samples (e.g., 48 variable positions over 2,500 bp) compared to both the African A/C subtypes of TMP-P with only eleven, or to only six across all eight tested Eurasian A/C subtype versions of TMP-P. Notably, all African A/C-like TMP-P genes (and the Saudi Arabian sample SKS1) have one common nucleotide change (431T) compared to all of the Eurasian versions, and therefore we shall refer to the African A/C-like group as the P[A/B1] subtype (Fig. 7).

Among the 23 examples sequenced of the African B subgroup of TMP-P allele genes (P[B] in Figs. 7 and 8), there were nine changes common to all of them. In addition, the predominant B1 subgroup (13 examples) all have six more common changes, whereas the B2 to B6 subgroups, which include eight of the nine examples from South Africa, lack those changes and are more variable. The six previously sequenced African TMP alleles (Kakoola et al. 2001) include three of our B1-subtype (Ugd2, Ugd12, Ugd19) and two of the A[B2] or A[B3] subset TMP-P alleles (Ugd4 and Ugd23) plus one B-subtype TMP-M allele (Ugd10) from Uganda.

5.3 Distinctive African Versions of TMP-M

Similarly, the seven sub-Saharan African TMP-M alleles that we have sequenced all proved to cluster into one of two groups, either those with typical B-subtype patterns similar to VG1 and Ug10 (5 examples) or those with typical Eurasian A/C-like patterns (2 examples). Therefore, although a subset of Eurasian TMP-P and TMP-M genes (like the A5 VIP genes) have seemingly returned from Eurasia to Africa, the majority of both are distinctive African

B subtypes that have not been found outside of either sub-Saharan Africa or related populations in America (e.g., Haiti and French Guyana). Interestingly, within African genomes there is little evidence for any linkage between VIP subtype genes (A5, B1, or B2) with any particular subtype of the TMP gene (whether $P_{A/C}$, P_B, $M_{A/C}$, or M_B), and examples of all combinations have been found (Figs. 7 and 8). Again, there was considerably more variation among the six known sequenced African TMP-M_B genes (six variable positions in all over 2,500 bp) compared to their invariant Eurasian TMP-$M_{A/C}$ counterparts, and SAKS26 and HKS56 both showed only one nucleotide difference from the five sequenced Eurasian TMP-$M_{A/C}$ genes. However, in addition, SAKS26 displays a 30-bp (10 amino acid) deletion within exon 8 (codon positions 389 to 398), and therefore represents the first and only example that we know about of a naturally deleted form of the TMP protein.

5.4 Distinctive Pacific Rim Versions of Both the TMP-P and TMP-M Alleles

We next considered whether, like the African specific versions of TMP-P and TMP-M, the KSHV genomes from the Pacific Rim branch might also show D/E subtype-specific TMP patterns. Although there were only four examples available in sufficient quantities to work with, namely, the two Hwalian samples from Taiwan of TKS10 and of TKS13 plus the two Polynesian ZKS3 and ZKS4 samples, these did indeed prove to have distinctive TMP-P(D) and TMP-M(E) patterns that are very different from those of any of the Eurasian A/C versions or the African B versions (Figs. 5 and 7). For the TMP-P(D) genes, TKS10, ZKS3, and ZKS4 were nearly identical across the entire 2,080-bp coding segment, but they displayed 43, 47, and 49 nucleotide differences (2.1%) from BCBL-R(A/C) and 33, 37, and 39 nucleotide differences (1.6%) from 431K(B), respectively. All three also included a 27-bp (9 amino acid) insertion after position 258 within exon 5. Similar to the B versions of TMP-P, TKS10, ZKS3, and ZKS4 also include a 10-bp insertion at the beginning of intron 1 compared to all A/C versions of TMP-P, but only TKS10, ZKS3, and ZKS4 also contain a 15-bp insertion in the noncoding region 93 bp upstream from the initiator codon. Out of 84 total variable nucleotide positions, the 431K(B1) and SAKS33(B2) variant TMP-P genes have nine differences from BCBL1(A/C) in common with TKS10(E) and ZKS4(D), but TKS10 and ZKS4 have 44 and 48 additional unique positions, respectively.

The TMP-M gene coding region from the Pacific Rim sample TKS13 shows 16 nucleotide differences from the prototype Eurasian version ASM70 (0.8%) and 22 from the prototype Africa version VG1 (1.1%). Out of a total of 33 variable nucleotide positions, HBL6(A/C) and TKS13(E) have 17 in common,

VG1(B) and TKS13(E) have 10 in common, and TKS13 has five unique changes. At the amino acid level, the 489-amino acid TMP-M proteins from HBL6(A/C) and TKS13(E) have six differences and VG1(B) and TKS13(E) have seven differences, whereas HBL6(A/C) and VG1(B) have 10 differences overall. VG1 and TKS13 have four amino acid differences from HBL6 in common, and TKS13 has two unique changes.

Therefore, the TKS13(E) subtype TMP-M gene parallels the VIP pattern by being intermediate between the HBL6(A/C) and VG(B) versions and more closely related to HBL6(A/C) than to VG1(B). In contrast, the TKS10 D-subtype TMP-P allele is instead much further diverged from BCBL1(A/C) than are the B versions, and it is 431K(B1) and SAKS33(B2) that are intermediate between TKS10(D) and BCBL1(A/C). For the TMP-P samples there are 25 amino acid differences between TKS10(D) and BCBL1(A/C) and 26 between TKS10(D) and 431K(B), compared to just five amino acid differences between BCBL1(A/C) and 431K(B1) or SAKS33(B2). Furthermore, among the TMP-P(D) genes, ZKS3 and ZKS4 differ from TKS10 at six and eight amino acid positions and between themselves at two positions. Therefore, the evolutionary distance of the TKS10/ZKS3/ZKS4(D) forms of the TMP-P protein appear to be between three and four times further diverged from both BCBL(A/C) and 431K(B1) or SAKS33(B2), compared to the distance that the TKS13(E) form of TMP-M has diverged from its BC1/HBL6(A/C) and VG1(B) counterparts.

5.5
Evidence for Three Distinct Chimeric Boundaries in African KSHV Genomes with TMP-M Alleles

We previously described a 500-bp transition region in HBL6 DNA encompassing the 3′-end of the TMP-M(K15) gene, a putative K14.1 gene, and the adjacent M-associated sequences at the 5′-end of the ORF75 gene (Poole et al. 1999). Within this region the level of sequence homology between the HBL6 M-prototype and the BCBL-R P-prototype changes gradually from 98% down to 80% and then abruptly becomes undetectable except for the coding sequences for several conserved SH3 and TRAF motifs in TMP. This clearly represents a complex ancient recombination junction between the modern genome constant region and a TMP-M allele of exotic origin. Our previously described diagnostic triple primer PCR test across the boundary here permits simple PCR product size-based discrimination between the standard P genomes and the less common P/M chimeric genomes (Poole et al. 1999).

Because of the patchwork homology over this region, the RHS of the chimeric domain is also readily detected within the sequence differences

chart for UPS75, one of our previously described constant region PCR loci (Zong et al. 1997). The distinctive exotic M pattern begins at position 1357 in UPS75 and extends rightward. Importantly, all 13 European and Asian M allele genomes analyzed in the UPS75 locus have essentially the same adjacent, M-associated UPS75 subtype K/M sequences on the LHS, which are very distinctive from all A/C-, B-, and D-subtype UPS75 sequences.

In contrast, among the nine African genomes with TMP-M alleles, there are three distinctly different chimeric patterns. Six samples, including all three American or Haitian genomes (VG1, OKS7, and OKS8) (Fig. 7), and three Ugandan genomes (HKS11, HKS41, and HKS49) (Fig. 8), proved to have an identical recombination junction boundary mapping between UPS75 positions 1220 and 1280 and producing a chimeric B3/M pattern. The Ugd10 sequence reported by Kakoola et al. (Kakoola et al., 2001) also has this same B3/M junction. In contrast, HKS56 displays a second different chimeric pattern referred to as B6/M, with a junction between positions 1290 and 1350, whereas in the two B/R/M′ chimeras found in SAKS26 (Fig. 7) and FTKS2 (not shown), the UPS75 sequence pattern (referred to as M′) is very similar throughout to the standard Eurasian K/M pattern, although it is distinctive at four positions. These latter also have chimeric B to R/M′ boundaries further to the left inside the adjacent UPS75E locus.

Consistent with these patterns, all seven B3/M genomes have the distinctive M[B] African B type of TMP-M gene, whereas both HKS56 and SAKS26 have novel M[A] versions of the Eurasian A/C-subtype TMP-M alleles. Therefore, we suggest that the B3/M genomes all have a common origin from a different source of TMP-M than the Eurasian versions, whereas SAKS26 (plus FTKS2) and HKS56 each arose by separate and more recent secondary recombination events with Eurasian-derived TMP-M containing genomes. Whether the African TMP-Ms once had a common origin with their Eurasian counterparts before the divergence of their migrating human hosts 100,000 years ago, or instead represent more modern independent recombination events from two different Neanderthal K/M-like rhadinovirus genomes, for example, remains an interesting speculation.

Finally, both TKS10 and TKS13 also display complex patterns in UPS75 that suggest possible additional chimeric junctions. In fact, all three extensively analyzed D or E genomes (TKS10, TKS13, and ZKS3) contain similar but distinctive K/M-like segments between UPS75 positions 950 and 1150 or 950 and 1130, with TKS13 also having an additional novel upstream region between positions 720 and 950.

6
The Rare KSHV N Alleles in Southern Africa

6.1
Additional N-Subtype Genomes at the UPS75 Locus

Alagiozoglou et al. (Alagiozoglou et al., 2000) first described another distinctive KSHV subtype within the 850-bp UPS75 locus from among a set of KS lesion DNAs from South African patients seen in Johannesburg. This set of six samples proved to be almost twice as far diverged from the BCBL-R A/C prototype as were even the M allele-associated K/M sequences in UPS75. We initially confirmed and extended these results with two of the same N-subtype samples described by Alagiozoglou et al. (SALN7, SALN8) (Zong et al. 2002) and have subsequently also found five more examples, three from South Africa (SAKS23, SAKS25, SAKS30) and two from Uganda (HKS35, WKS14). The N-subtype sequence for the UPS75 region between positions 540 to 1300 has 42 nucleotide differences (5.5%) from the UPS75 P(A/C) prototype pattern of BCBL-R, although they vary at just one position among themselves. This compares to 16 differences in this region (2.1%) between BCBL-R and the Eurasian M-subtypes such as ASM70, which again show only one variant position among all 15 European, Asian, and African examples analyzed (Poole et al. 1999). By comparison, there are usually a total of only seven to nine nucleotide differences here from BCBL-R in each of the several different variants of African B-subtype UPS75 genes found. Finally, at this locus, the D1 and D2 subtypes found in ZKS3, ZKS4, and TKS10 were identical, but they all differ at 17 positions from BCBL-R, whereas TKS13(E2) is closely related to the Ds but is also more complex (22 differences from BCBL-R). The branching and divergence relationships among various different UPS75 (as well as LANA and ORF-75E) subtypes are illustrated in Fig. 9.

6.2
Novel N-Allele KSHV TMP Genes from Southern Africa

Obviously, the N-subtype UPS75 data suggested that the TMP genes from these KSHV samples might be very different from either the B-subtype TMP-P or B-subtype TMP-M alleles. Indeed, this proved to be correct for all four different KS lesions studied (SAKS23, SAKS25, SAKS30, and HKS35) that had N-subtype UPS75 genes and were available for further analysis. Two of these samples (SAKS25 and SAKS30) were derived a year apart from the same patient, but all four proved to be identical over the entire length of their TMP-N genes. Initially, we were unable to detect any PCR products directly with the standard triple primer reaction from this particular subset of

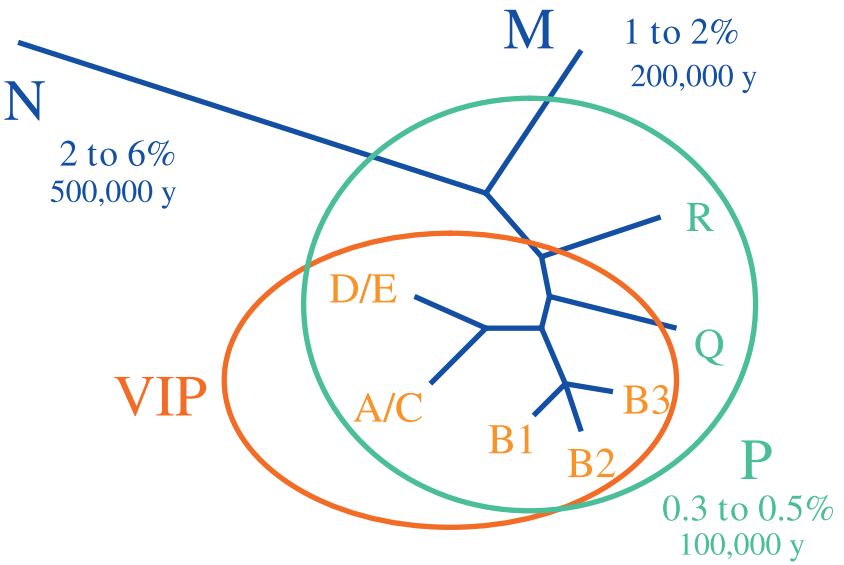


Fig. 9 Summary diagram illustrating the interpreted relative phylogenic ages for the subtype divergence patterns found within the RHS constant region segments in the LANA, 75E, and UPS75 loci. Note that in addition to the A/C, D/E, B1/B2/B3, Q, and R branches, which are all designated as the P-allele subset (encompassed by the *yellow circle*); there are also older more diverged M and N allele-associated segments with 1% to 2% (M) and 2% to 6% (N) nucleotide level variation from the PA/C pattern. The *red oval* encompasses the "modern human" subset of the P-allele VIP gene patterns (A/C, D/E, and B), which are the only subtypes now found at the extreme LHS of KSHV genomes

African samples that were known to have N-subtype UPS75 genes. They were still negative even after hybridization to Southern blots of the PCR DNA with either specific ^{32}P-labeled UPS75-P or ^{32}P-labeled UPS75-M DNA probes (data not shown). However, PCR products were ultimately recovered and sequenced from the SAKS23 prototype-N sample by using two rounds of redundant RHS 5′-primers whose design was based on weak residual DNA homology near the initiator codons of both TMP-P and TMP-M. Subsequently, a series of eight new TMP-N specific primers were generated and used to successfully amplify and sequence the entire 2,500-bp TMP gene and surrounding region from all four previously negative samples suspected of being N subtypes.

This new subset of KSHV TMP genes proved to fall into an intermediate position between the TMP-P and TMP-M allele patterns, although being considerably closer to the TMP-P pattern (Fig. 5). At the nucleotide level, the 2,080-bp SAKS23 coding region DNA including the introns shows 17% differ-

ences from the BCBL-R gene, whereas at the protein level, the novel N-associated TMP gene displays 29% amino acid differences from the prototype TMP-P protein (BCBL-R) and 69% differences from the prototype TMP-M protein (HBL6). Obviously, this is vastly greater than the 0.5% to 2% nucleotide differences for B and D subtypes relative to the prototype A/C subtypes for both TMP-P and TMP-M. Therefore, we consider that TMP-N represents a third distinctive allele with evolutionary hierarchal stature equivalent to that of the TMP-P and TMP-M alleles themselves.

Interestingly, the major structural difference in the TMP-N protein compared to both TMP-P and TMP-M is the insertion of an extra nine amino acids after codon 258 in exon 5, which lies within the fifth extracellular loop. Remarkably, this is also almost exactly at the same location as the novel nine-amino acid insert within the subtype-D TMP-P gene (TKS10, ZKS3, and ZKS4). Furthermore, there is nucleotide homology at 20 out of 27 bp between these two inserts (75% identity), and both encode an NXT/S *N* -glycosylation motif at exactly the same position. This similarity indicates that the extra domain evidently must represent the primordial situation, with both the M and all other P samples having suffered independent deletions of this NXT/S motif. Each TMP allele has only one other potential NXT/S motif, which maps within extracellular loop 2 for TMP-P and TMP-N and within extracellular loop 3 for TMP-M.

Overall, the 12 transmembrane domains of the three TMP allele proteins show considerably less divergence than the intervening loop domains, and, most importantly, several blocks of amino acids within the cytoplasmic tail of TMP (encoded entirely by exon 8) are virtually completely conserved in all three versions, including regions implicated previously as SH2 or SH3 tyrosine kinase interaction motifs such as YEEVL and YASIL, or the two potential TRAF interaction motifs QSGIS and TQPTD, although a third, SPQPD, is absent in TMP-M (Glenn et al. 1999). Curiously, like both the subtype-B and -D versions of TMP-P, TMP-N also contains a 10-bp insertion at the beginning of intron 1 relative to the TMP-$P_{A/C}$ version, and the sequence here is only 2 bp different from that in the TMP-P_B versions. Two internal 23-bp and 43-bp segments of TMP-N introns 1 and 2 are also very different from the TMP-P versions (14- and 27-bp deletions or mismatches, respectively), and there are also five other 1-bp insertions or deletions spread across introns 4, 5, 6, and 7.

6.3
Extension into the ORF75E Locus Reveals Additional RHS Chimeric Junctions

Data for three of the N-subtype samples (SAKS23, SAKS25, and HKS35) were extended to cover all of the adjacent expanded ORF75E locus (Chang

et al. 1994; Zong et al. 1997), where they differ at 18 polymorphic positions from the BCBL-R (A/C) prototype, although in this segment they also more closely resemble several other patterns, including K/M (4 differences), Q (4 differences), and R (5 differences). In comparison, the D, B, and F patterns for ORF75E differ from A/C at only 6, 9, and 8 positions, respectively. One important additional sample ZKS6 from a Mozambique San bushman patient was also judged to be N subtype at this locus (Fig. 7).

At least three additional chimeric boundaries are detectable within the ORF75E locus in other African genomes. The most common one occurs between positions 159 and 237 in 10 samples (nine from Uganda) that we refer to as the R/A subtype (Fig. 8). All were A[B] in ORF75 and also have common patterns in T0.7 and LANA. Eight were sequenced for TMP, and all were judged to be P[AB1]. Therefore, the size of the RHS Eurasian-derived A/C-like segment present in nearly half of the African chimeric genomes is only 2.6 kb. Interestingly, the second type of chimera here (R/M) also involves almost exactly the same location, but this time joining an R pattern on the left with a K/M pattern on the right in SAKS26 (Fig. 7) and HKS67 (not shown). The third obvious chimeric junction in this region involves an A/C-like pattern to the left of position 528 joined to a B3 pattern on the right in all four American B genomes (JKS15, OKS7 and OKS8, and VG1) (Fig. 7). HKS54 and HKS56 are also very similar, but with slightly different boundaries (A/G and A/B6) (Fig. 8).

7
Differential Evolutionary History of Segments Along the KSHV Genome

7.1
Age of Divergence and Origin of the Three TMP Alleles

In attempting to estimate the age of the evolutionary divergence of TMP-P, TMP-N, and TMP-M alleles from each other, we have based our calculations on the presumption that the B subtypes of both TMP-P and TMP-M (which show between 0.9% and 1.2% nucleotide divergence each from the A/C forms) are judged to have branched away from their common modern human ancestor close to 100,000 years ago. Intriguingly, although divergence of the D/E (Pacific Rim) version of TMP-M in TKS13 (0.85%) fits well with the notion of being 70,000 years separated from the A/C and B versions, the particular TMP-P allele present in TKS10, ZKS3, and ZKS4 (2.3% diverged) is obviously much older than the B version and is projected to have diverged from both the A/C and B versions between 200,000 and 250,000 years ago (Figs. 1 and 5). An

extrapolation for 21.3% nucleotide divergence would suggest an evolutionary distance of 2 to 4 million years for the TMP-N allele from TMP-P, with a further extrapolation to at least 10 to 15 million years for TMP-M from TMP-P. Note that the equivalent gene (R15) in the RRV rhesus rhadinovirus (the RV2 prototype herpesvirus), although showing a strong structural resemblance with 12 transmembrane domains and eight matching coding exons, has no measurable residual amino acid homology to any of these KSHV TMP alleles (Alexander et al. 2000; Searles et al. 1999). The original divergence of rhesus monkeys from great apes and humans is expected to be on the order of 20 to 25 million years ago, but because KSHV fits into the RV1 rather than RV2 lineage of Old World primate rhadinoviruses (Rose 2003; Rose et al. 1997), the RRV difference is probably misleadingly large and that in rhesus RFMHV (an RV1 class) virus would likely provide a better yardstick. In the future, sequence data for the TMP genes from the three distinct KSHV-like viruses known in chimpanzees (Lacoste et al. 2000c), which would be expected to have diverged from the equivalent human and prehuman versions only 5 to 6 million years ago, will go a long way toward clarifying the exact age of divergence of the three TMP alleles found in modern humans.

Whether the events that generated these chimeras occurred between multiple distinct viruses all infecting the same prehuman hosts or instead were derived by recombination with exotic nonhuman but nevertheless "hominid" lineages remains to be elucidated. However, it is very clear from the easily recognized recombination junctions found near the N-terminal regions of ORF75 that both the TMP-N and TMP-M alleles themselves are the only segments of modern-day KSHV genomes that diverged more than 1 million years ago, whereas all of the genomes carrying them that have been detected so far represent chimeras containing these relics of anciently diverged viruses recombined into predominantly modern lineage human KSHV genomes. Nevertheless, the modern KSHV genomes that they originally recombined with (as represented by the adjacent constant regions in the T0.7, LANA, ORF75E, and UPS75 loci, etc.) are also both very distinctive from and significantly older than the standard subtype A/C, B, and D/E level divergence seen in the VIP gene at the LHS and in the central constant segment loci of all other sequenced KSHV genomes. Recombination with, for example, *Homo neanderthalis*, *Homo erectus*, or *Homo robustus* versions of KSHV would be the type of scenario that we envisage may have introduced these small segments of now presumably extinct KSHV-like viruses into chimeric modern human genomes.

However, we cannot exclude the alternative possibility that the complete highly diverged original M genomes that evolved separately in prehumans may still exist in modern *Homo sapiens* but have not yet been detected, perhaps because they are not associated with KS. Consider that (1) chimpanzees are

known to have three distinct rhadinoviruses, two in the RV2 RRV-like KSHV-like lineage and one in the RV1 lineage (Lacoste et al. 2001); (2) all three major branches of modern humans (Eurasian, African, and south Asia-Pacific Rim) have their own distinctive versions of both the hugely diverged TMP-P and TMP-M alleles as either complete P or chimeric P/M KSHV genomes; and (3) at least one of the oldest groups of humans in southern Africa has a representative of yet a third rhadinovirus TMP allele as P/N chimeras. Therefore, the possibility that there are or were three distinct human RV1 lineage rhadinovirus species cannot be ignored.

7.2 Timing of the P/M Chimeric Events

Under either scenario, a question also arises about whether the final recombination event(s) that created the P/M chimeras occurred just once before modern humans came out of Africa or occurred on three separate occasions, once in the African, once in the Pacific Rim, and once in the Eurasian lineage. For the first option to be correct, the small founding human population that initially successfully exited the African continent would have had to have brought two viruses with it, one P and the other P/M, both of which then independently diverged within the three main branches. However, the fact that the three different chimeras each have different junction boundaries clearly favors the alternative scenario in which just a single *Homo sapiens* P genome came out of Africa along with the founding modern humans, but that each of the three branches independently recombined with pre-existing Neanderthal version P/M chimeras that had also diverged into three separate lineages in the three different continental subgroups. In fact, we probably also have to conclude that this type of event occurred at least twice in Africa to create both the relatively common B3/M chimera and the rare B6/M chimera found in HKS56. In addition, a secondary transfer event must have also introduced the Eurasian type M′[A] exotic TMP allele into a B/R genome in Africa (SAKS26 and FTKS2).

We originally favored the scenario that the source (in terms of both time and place) for the origin of the KSHV P/M chimeric genomes was the still overlapping *Homo sapiens* and Neanderthal populations in the Middle East as recently as 25,000 years ago, but the new evidence both here and from Kakoola et al. (Kakoola et al. 2001) for distinct African and now also south Asia-Pacific Rim versions of both the TMP-P and TMP-M alleles instead implies an alternative model in which the original TMP-M allele (probably already in the form of a P/M chimera) must have already been present before the migrations out of Africa. However, whether it was present within modern

Homo sapiens at that time or rather just in an older "humanoid" subgroup is a critical point. If the former, this second type of KSHV virus would have had to be retained in all three branches during the same founder bottleneck events that also led to the divergence of the P versions of the VIP gene into just one main subtype (A/C, B, or D/ E) within each branch. This seems doubtful. Therefore, a more complex version of the original model suggested above, in which all three subsets of TMP-M alleles were instead acquired independently by multiple but rare transfers into modern human P genomes from already existing older and already diverged "Neanderthal" chimeric P/M genomes that occurred after the primary modern *Homo sapiens* migrations out of Africa, seems to be a much more likely prospect. This model also suggests that some form of selective advantage for these particular chimeric recombination products may have been involved (within modern humans at least). Furthermore, the unexpectedly high divergence of the south Asian-Pacific Rim D version of TMP-P, but not of the E version of TMP-M, provides an additional complication to the single parental "out of Africa" concept that needs to be considered in any all encompassing model. However, it could be entirely consistent with divergence of the parental "Neanderthal" P/M genomes into three branches with a temporal pattern different from that for modern humans.

7.3 Source of the Exotic TMP-N Alleles

Presumably the rare P/N genomes would have to represent another example of a similar scenario in Africa, although presumably in this case with a different prehuman KSHV-like virus having a TMP gene allele that had diverged at an intermediate temporal stage between the P and M allele versions. However, it is still necessary to evaluate the full length of the central segments of all of these genomes more thoroughly before any compelling theories can be developed about the source, age, origin, and history of the original parental N-class KSHV genome. Nevertheless, it is probably highly significant that four of the five examples of genomes with both N-subtype TMP alleles and large N-associated adjacent internal segments came from South African sources (and the other one was from Uganda). We have also observed a number of additional South African KS genomes (but rarely Ugandan ones) that have partial N-like constant region segments, but without the associated N-allele TMP genes (Zong J-C and Hayward, G.S., unpublished data). Note that many of these South African patients came from a mining area where there is significant mixed heritage involving Zhosa or Khoisan ancestry. Furthermore, ZKS6, which contains an N-subtype ORF75E gene at least, came from a patient of

known San bushman ancestry who emigrated to South Africa and then to New Zealand from Mozambique. Similarly, Whitby et al. (Whitby et al., 2004) have recently described a San bushman KSHV sample from Botswana (SAN1) that also contains an N-subtype ORF75 gene. These observations strongly hint that both the TMP-N allele and the N-subtype ORF75 gene may be indigenous in Zhosa or Khoisan populations (and may never have left Africa).

7.4 Only Modern Versions of the VIP Genes Have Been Detected

In the case of the VIP genes on the LHS of modern human KSHV genomes, there seems to be a very different evolutionary dynamic in action. Two overall general conclusions can be made. First, only modern P forms of the VIP gene are found, which split relatively recently into the A/C, B, and D/E branches corresponding to the diasporas created by migrations out of East Africa. Second, in sub-Saharan Africa, the VIP genes are essentially randomized and unlinked relative to their associated adjacent constant region subgroup patterns. In distinct contrast to the retention of alternative TMP alleles that are ancient chimeric relics of other viruses on the RHS, no VIP gene subtypes older than the estimated 100,000 years divergence created by the migrations into and out of Africa have been detected. Admittedly, we probably would not be able to detect true M or N versions of the VIP gene directly with the P-subtype primers that we use. However, no KSHV genomes that are detectable with ORF26, T0.7, or UPS75 locus primers (which do amplify all N, M, B, R, and Q constant region variants) have ever proved not to have a detectable linked modern P-subtype VIP gene. Furthermore, all VIP genes found so far fit into the B, A/C, D/E, or F/G branches of the P subtype. Therefore, on the basis of the temporal yardstick of 40,000 years divergence from each other for the known aboriginal Hwalien (Taiwan), Australian, Polynesian, Japanese (Ainu), and Amerindian versions within the D/E branch, we estimate that all known versions of VIP diverged no more than 100,000 years ago. We envisage that the B branch separated from the common ancestor of all of the others approximately 100,000 years ago, then the D/E branch diverged from the A/C and F/G line 70,000 years ago, and finally the A/C branch diverged from the F/G branch about 40,000 to 50,000 years ago. Subsequently, the Eurasian A plus C branches likely diverged from each other no more than 30,000 years ago, with the four main subclusters (C′, C″, A′, A″) each having branched off 15,000 years ago. Similarly, in sub-Saharan Africa, the original divergence of all known B-subtype VIP genes into the B1 and B2 variants appears to have occurred between 25,000 and 30,000 years ago, but in contrast the entire A5 cluster evidently originated no more than 4,000 years ago.

7.5 Source and Rapid Penetrance of the Anomalous A5 VIP Subtype Within Sub-Saharan Africa

How then did the "European-like" A5 VIP genes become distributed seemingly uniformly throughout central, eastern and southern Africa, as well as become associated almost randomly with genomes that contain B, R, Q, or N constant regions? The interpretation partially depends on the age of divergence of the constant region subtypes themselves, which is very difficult to estimate and clearly varies between gene loci with a pronounced tendency to be greater toward the RHS than on the LHS (Zong, J-C and Hayward, G.S., unpublished data). At nearly all constant region loci, we recognize five distinctive African subtype clusters (referred to as B, R, Q, N, and M′) that are at least as far diverged from the A/C cluster as are the Bs, as well as F-, G-, D/E-, J-, and K-subtype clusters that are intermediate between the A/C and B subtypes. Notably, B, R, Q, and N are all unique to sub-Saharan Africa. Consequently, we judge that the R, Q, N, and M subtypes of constant region sequences on the RHS all diverged from a common ancestor with the A/C and B subtypes well before the "out of Africa" migrations, whereas on the LHS some may have diverged contemporaneously with B from A/C. Overall, it is very obvious that there have been no VIP genes yet detected that have the equivalent patterns of divergence expected for the older R, Q, N, and M versions of VIP. Within sub-Saharan Africa, they have all evidently been lost and replaced by relatively modern B1, B2, or A5 VIP genes.

Evidently, a prototype A5 VIP gene must have been originally introduced from Europe or north Africa (probably just once) into presumably a B-subtype genome in the ancestors of the Bantu people at Bok in Nigeria before their expansion across most of the continent beginning about 4,000 years ago. Recombination into the genomes of other ethnic African populations who were displaced or absorbed would account for some spread into the hybrid B/R and B/Q genomes that are now most common among the predominantly Bantu samples that we analyzed in Uganda. However, the Bantu expansion did not proceed beyond the Cook River in South Africa before Europeans arrived several hundred years ago. Yet many of the South African genomes with N-subtype segments, which are presumably of Zhosa or Khoisan origin, also have A5 VIP genes! Furthermore, even the B-subtype VIP genes from South Africa are very little different from those found in Uganda or elsewhere in Africa, in contrast to the often dramatic differences in their constant region subtype patterns (i.e., including Q, R, and N subtypes). We propose that the most likely explanation is a very rapid, and recent, "aggressive" spread through recombination events, which are also presumably connected to some strong

selective advantage in combining a Eurasian A5-subtype VIP gene with otherwise African-specific B-, Q-, R-, or N-subtype constant region genomes. Unfortunately, there is too little divergence in the ORF4 region adjacent to VIP to be able to detect chimeric boundaries there or to judge whether or not the A5/B LHS segment was subsequently introduced into R, Q, and N genomes on just one or on multiple occasions (although we assume the latter).

8
Overall Significance

8.1
Precedents From Other Herpesviruses

There are several obvious precedents for multiple distinct herpesvirus species from the same virus genus coexisting within the same host species. For example, human HSV-1 and HSV-2 have an average protein level divergence of approximately 30%, which is similar to the average protein level divergence of human and chimpanzee CMVs (Davison et al. 2003). HSV-1 and HSV-2 have different preferential sites for establishment of latency in their human hosts and evidently do not successfully recombine in vivo, but can do so relatively easily in coinfected cell cultures under laboratory conditions. Similarly, humans harbor three distinct roseoloviruses with similar but distinct biological properties, HHV6A, HHV6B, and HHV7. The overall divergence level of HHV6 and HHV7 is similar to that between HSV-1 and HSV-2, or between human and chimpanzee CMVs, but although the level of divergence of several genes in HHV6A and HHV6B reaches 15% to 25% at the protein level, most of the rest of their genomes differ by no more than 5% to 10% at the nucleotide level. Again, no known chimeras have been detected, but most individuals harbor HHV7 plus one or both of the HHV6 variants simultaneously.

Finally, the human lymphocryptovirus EBV has genomes that fall into two subtypes, often referred to as EBV-1 and EBV-2 or, preferably, EBV-A and EBV-B. Here the situation more closely resembles that in KSHV, with one gene in particular, EBNA2, occurring as two highly dimorphic alleles known as EBNA2A and EBNA2B that differ by approximately 45% at the amino acid level (Dambaugh et al. 1984), compared to a 55% amino acid divergence for both from the equivalent gene in the baboon lymphocryptovirus, H. papio or HPV1 (Ling et al. 1993). Several of the other EBV latency genes also split into A and B subtypes but show only 3% to 5% amino acid divergence, and, except for the somewhat hypervariable LMP1 gene, the differences elsewhere are likely to be no more than a few percent at the nucleotide level.

Therefore, similar to the situation with the KSHV TMP alleles, one of the two forms of human EBV genomes evidently carries a chimeric segment consisting of just the second allelic type of EBNA-2 gene, which must have diverged perhaps 15 to 20 million years ago, and would seem likely to have been derived by an exotic cross-species recombination event into a modern human EBV genome that was slightly different from the current predominant version of human EBV-A. Just as with KSHV, some relatively rare recombinant chimeras containing segments of both EBV-A and EBV-B are known, but the two EBNA2 alleles are both found spread throughout the Eurasian, African, and south Asia-Pacific Rim (Papua New Guinea) branches of modern humans (Aitken et al. 1994). There also appear to be distinctive Eurasian, African, and south Asian-Pacific versions of at least the EBNA2A allele (Shim et al. 1998).

Although it seems extremely unlikely that a complete version of the parental EBV-B-like virus that was the source of the second human EBNA2 allele still exists in modern humans, it does seem plausible that either a complete or a chimeric version of it may have existed in some other now extinct branch of prehumans, such as Neanderthals, and was then transferred probably only once early during the period when modern humans overlapped with them, originally perhaps 100,000 to 200,000 years ago in Africa or even as recently as 25,000 to 30,000 years ago within the Middle East or southern Europe.

8.2 Contribution of Chimeric Fixation to Herpesvirus Genomic Evolution

Irrespective of their precise origins, the presence of these chimeric genomes and alternative alleles of key genes in both KSHV and EBV probably reflect a somewhat unexpected but normal mechanism by which herpesvirus genomes evolve. There are also three radically different alleles of the STP/TIP genes in herpesvirus saimiri (HVS) that may represent a similar situation. Even although herpesviruses are not supposed to successfully cross between species barriers as whole genomes, our data here provide evidence that partial chimeric versions of exotic viral cross-species recombination events probably do occur and do survive successfully at least temporarily (irrespective of the issue of whether a cross-species host transfer is also involved). Presumably, for closely related viral sequences with greater than 85% nucleotide identity, unless there is a selective advantage, the rare novel form will tend to be diluted out and lost by homologous recombination, as we see occurring with the progressively shorter segments of the M-associated K/M constant region sequences found in Eurasian KSHV genomes that retain TMP-M (Kakoola et al. 2001; Poole et al. 1999; Zong et al. 2002). However, for genes that are sufficiently diverged already to be unable to undergo homologous recombi-

nation, that is, greater than 15% differences at the nucleotide level, such as TMP-P, -M, and -N or EBNA2A and 2B, they would not become scrambled or diluted out, especially if there is a positive selective biological effect, and can become fixed in the population, initially just as alternative alleles but eventually perhaps after certain bottlenecks or founder effects narrow down the range of variants in a subpopulation, they could even be destined to take over as the only extant version for a particular virus species.

Acknowledgements These studies were supported by NIH Research Grants R01-CA-73585 and P01-CA-81400 awarded to G.S.H. We thank the many colleagues who provided anonymous DNA samples or clinical paraffin sections for analysis that made this work possible. Many of them were acknowledged previously (32, 44, 45), but we especially want to thank our colleagues Henry Kajumbula, Edward Katange-Mbide, and William Boto from the Uganda Cancer Institute, University of Kampala, Uganda, and Lee Alagiozioglou, Lynn Moris, and Fred Sitas from The National Institute of Virology, Sandringhan, South Africa, for the Ugandan HKS and South African SAKS collections, respectively, for which the primary data reports have not yet been published. We also acknowledge the major contribution of Peter S. Kim to the identification of the TMP-N allele.

References

Aitken C, Sengupta SK, Aedes C, Moss DJ, and Sculley TB (1994). Heterogeneity within the Epstein-Barr virus nuclear antigen 2 gene in different strains of Epstein-Barr virus. J Gen Virol *75*, 95–100

Alagiozoglou L, Sitas F, and Morris L (2000). Phylogenetic analysis of HHV-8 in South Africa and identification of a novel subgroup. J Gen Virol *81*, 2029–2038

Alexander L, Denekamp L, Knapp A, Auerbach MR, Damania B, and Desrosiers RC (2000). The primary sequence of rhesus monkey rhadinovirus isolate 26–95: sequence similarities to Kaposi's sarcoma-associated herpesvirus and rhesus monkey rhadinovirus isolate 17577. J Virol *74*, 3388–3398

Biggar R, Whitby D, Marshall V, Linhares AC, and Black F (2000). Human Herpesvirus 8 in Brazilian Amerindians: A hyperendemic population with a new subtype. J Infect Dis *181*, 1562–1568

Brinkmann MM, Glenn M, Rainbow L, Kieser A, Henke-Gendo C, and Schulz TF (2003). Activation of mitogen-activated protein kinase and NF-κB pathways by a Kaposi's sarcoma-associated herpesvirus K15 membrane protein. J Virol *77*, 9346–9358

Chang Y, Cesarman E, Pessin MS, Lee F, Culpepper J, Knowles DM, and Moore PS (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science *266*, 1865–1869

Choi J-K, Lee B-S, Shim SN, Li M, and Jung JU (2000). Identification of the novel K15 gene at the rightmost end of the Kaposi's sarcoma-associated herpesvirus genome. J Virol *74*, 436–446

Chokunonga E, Levy LM, Bassett MT, Borok MZ, Mauchaza BG, Chirenje MZ, and Parkin DM (1999). AIDS and cancer in Africa: the evolving epidemic in Zimbabwe. AIDS *13*, 2583–2588

Cook PM, Whitby D, Calabro ML, Luppi M, Kakoola DN, Hjalgrim H, Ariyoshi K, Ensoli B, Davison AJ, and Schulz TF (1999). Variability and evolution of Kaposi's sarcoma-associated herpesvirus in Europe and Africa. International collaborative group. AIDS, 1165–1176

Damania B, and Jung JU (2001). Comparative analysis of the transforming mechanisms of Epstein-Barr virus, Kaposi's sarcoma-associated herpesvirus, and *Herpesvirus saimiri*. Adv Cancer Res *80*, 51–82

Dambaugh T, Hennessey K, Chamnankit L, and Kieff E (1984). U2 region of Epstein-Barr virus DNA may encode Epstein-Barr virus nuclear antigen 2. Proc Natl Acad Sci USA *81*, 7632–7636

Davison AJ, Dolan A, Akter P, Addison C, Dargan DJ, Alcendor DJ, McGeoch DJ, and Hayward GS (2003). The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. J Gen Virol *84*, 17–28

Dedicoat M, and Newton R (2003). Review of the distribution of Kaposi's sarcoma-associated herpesvirus (KSHV) in Africa in relation to the incidence of Kaposi's sarcoma. Br J Cancer *88*, 1–3

Fouchard N, Lacoste V, Couppie P, Develoux M, Mauclere P, Michel P, Herve V, Pradinaud R, Bestetti G, Huerre M, et al. (2000). Detection and genetic polymorphism of human herpesvirus Type 8 in endemic or epidemic Kaposi's Sarcoma from West and Central Africa, and South America. Int J Cancer *85*, 166–170

Glenn M, Rainbow L, Aurade F, Davison A, and 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

Hayward GS (1999). KSHV Strains: The origin and global spread of the virus. (Review) Semin Cancer Biol, 187–199

Kadyrova E, Lacoste V, Duprez R, Pozharissky K, Molochkov V, Huerre M, Gurtsevitch V, and Gessain A (2003). Molecular epidemiology of Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8 strains from Russian patients with classic, posttransplant, and AIDS-associated Kaposi's sarcoma. J Med Virol *71*, 548–556

Kajumbula H, Wallace RG, Zong JC, Hokello J, Sussman N, Simms S, Rockwell RF, Pozos R, Hayward GS, and Boto W (2006). Ugandan Kaposi's sarcoma-associated herpesvirus phylogeny: Evidence for cross-ethnic transmission of viral subtypes. Intervirology *49*, 133–143

Kakoola DN, Sheldon J, Byabazaire N, Bowden RJ, Katongole-Mbidde E, Schulz TF, and Davison AJ (2001). Recombination in human herpesvirus-8 strains from Uganda and evolution of the K15 gene. J Gen Virol *82*, 2393–2404

Kasolo FC, Mpabalwani E, and Gompels UA (1997). Infection with AIDS-related herpesviruses in human immunodeficiency virus-negative infants and endemic childhood Kaposi's sarcoma in Africa. J Gen Virol *78*, 847–856

Lacoste V, Judde JG, Briere J, Tulliez M, Garin B, Kassa-Kelembho E, Morvan J, Couppie P, Clyti E, Forteza Vila J, *et al.* (2000a). 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

Lacoste V, Kadyrova E, Chistiakova I, Gurtsevitch V, Judde JG, and Gessain A (2000b). Molecular characterization of Kaposi's sarcoma-associated herpesvirus/human herpesvirus-8 strains from Russia. J Gen Virol *81* Pt 5, 1217–1222

Lacoste V, Mauclere P, Dubreuil G, Lewis J, Georges-Courbot MC, and Gessain A (2000c). KSHV-like herpesviruses in chimps and gorillas. Nature *407*, 151–152

Lacoste V, Mauclere P, Dubreuil G, Lewis J, Georges-Courbot MC, and Gessain A (2001). A novel gamma 2-herpesvirus of the Rhadinovirus 2 lineage in chimpanzees. Genome Res *11*, 1511–1519

Lagunoff M, and Ganem D (1997). The structure and coding organization of the genomic termini of Kaposi's Sarcoma-associated herpesvirus (Human Herpesvirus 8). Virology *236*, 147–154

Lee BS, Connole M, Tang Z, Harris NL, and Jung JU (2003). Structural analysis of the Kaposi's sarcoma-associated herpesvirus K1 protein. J Virol *77*, 8072–8086

Lee BS, Lee SH, Feng P, Chang H, Cho NH, and Jung JU (2005). Characterization of the Kaposi's sarcoma-associated herpesvirus K1 signalosome. J Virol *79*, 12173–12184

Ling PD, Ryon JJ, and Hayward SD (1993). EBNA-2 of herpesvirus papio diverges significantly from the type A and type B EBNA-2 proteins of Epstein-Barr virus but retains an efficient transactivation domain with a conserved hydrophobic motif. J Virol *67*, 2990–3003

Meng YX, Sata T, Stamey FR, Voevodin A, Katano H, Koizumi H, Deleon M, De Cristofano MA, Galimberti R, and Pellett PE (2001). Molecular characterization of strains of human herpesvirus 8 from Japan, Argentina and Kuwait. J Gen Virol *82*, 499–506

Meng YX, Spira TJ, Bhat GJ, Birch CJ, Druce JD, Edlin BR, Edwards R, Gunthel C, Newton R, Stamey FR, *et al.* (1999). Individuals from North America, Australasia, and Africa are infected with four different genotypes of human herpesvirus 8. Virology *261*, 106–119

Nicholas J, Zong J-C, Alcendor DJ, Ciufo DM, Poole LJ, Sarisky RT, Chiou CJ, Zhang X, Wan X, Guo H-G, *et al.* (1998). Novel organizational features, captured cellular genes and strain variability within the genome of KSHV/HHV8. J Natl Cancer Inst Monogr *23*, 79–88

Poole LJ, Zong J-C, Ciufo DM, Alcendor DJ, Cannon JS, Ambinder R, Orenstein J, Reitz MS, and Hayward GS (1999). Comparison of genetic variability at multiple loci across the genomes of the major subgroups of Kaposi's sarcoma associated herpesvirus (HHV8) reveals evidence for recombination and for two distinct types of ORF-K15 alleles at the right hand end. J Virol *73*, 6646–6660

Rose TM, Ryan, J. T., Schultz, E.R., Raden, B.W., Tsai, C-C (2003). Analysis of 4.3 kilobases of divergent locus B of macaque retroperitoneal fibromatosis-associated herpesvirus reveals a close similarity in gene sequence and genome organization to Kaposi's sarcoma-associated herpesvirus. JVirol *77*, 5084–5097

Rose TM, Strand KB, Schultz ER, Schaefer G, G.W. Rankin J, Thouless ME, Tsai C-C, and Bosch ML (1997). Identification of two homologs of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in retroperitoneal fibromatosis of different macaque species. J Virol *71*, 4138–4144

Russo JJ, Bohenzky RA, Chien M-C, Chen J, Yan M, Maddalena D, Parry JP, Peruzzi D, Edelman IS, Chang Y, and Moore PS (1996). Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Proc Natl Acad Sci USA *93*, 14862–14867

Searles RP, Bergquam EP, Axthelm MK, and Wong SW (1999). Sequence and genomic analysis of a rhesus macaque rhadinovirus with similarity to Kaposi's sarcoma-associated herpesvirus/human herpesvirus 8. J Virol *73*, 3040–3053

Sharp TV, Wang HW, Koumi A, Hollyman D, Endo Y, Ye H, Du MQ, and Boshoff C (2002). K15 protein of Kaposi's sarcoma-associated herpesvirus is latently expressed and binds to HAX-1, a protein with antiapoptotic function. J Virol *76*, 802–816

Shim YS, Kim CW, and Lee WK (1998). Sequence variation of EBNA2 of Epstein-Barr virus isolates from Korea. Mol Cells *8*, 226–232

Stebbing J, Bourboulia D, Johnson M, Henderson S, Williams I, Wilder N, Tyrer M, Youle M, Imami N, Kobu T, *et al.* (2003). Kaposi's sarcoma-associated herpesvirus cytotoxic T lymphocytes recognize and target Darwinian positively selected autologous K1 epitopes. J Virol *77*, 4306–4314

Treurnicht FK, Engelbrecht S, Taylor MB, Schneider JW, and van Rensburg EJ (2002). HHV-8 subtypes in South Africa: identification of a case suggesting a novel B variant. J Med Virol *66*, 235–240

Wang L, Wakisaka N, Tomlinson CC, DeWire SM, Krall S, Pagano JS, and Damania B (2004). The Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) K1 protein induces expression of angiogenic and invasion factors. Cancer Res *64*, 2774–2781

Whitby D, Marshall VA, Bagni RK, Wang CD, Gamache CJ, Guzman JR, Kron M, Ebbesen P, and Biggar RJ (2004). Genotypic characterization of Kaposi's sarcoma-associated herpesvirus in asymptomatic infected subjects from isolated populations. J Gen Virol *85*, 155–163

Zhang YJ, Davis TL, Wang XP, Deng JH, Baillargeon J, Yeh IT, Jenson HB, and Gao SJ (2001). Distinct distribution of rare US genotypes of Kaposi's sarcoma- associated herpesvirus (KSHV) in South Texas: implications for KSHV epidemiology. J Infect Dis *183*, 125–129

Zong J-C, Ciufo DM, Alcendor DJ, Wan X, Nicholas J, Browning P, Rady P, Tyring SK, Orenstein J, Rabkin C, *et al.* (1999). High level variability in the ORF-K1 membrane protein gene at the left end of the Kaposi's sarcoma associated herpesvirus (HHV8) genome defines four major virus subtypes and multiple clades in different human populations. J Virol *73*, 4156–4170

Zong JC, Ciufo DM, Viscidi R, Alagiozoglou L, Tyring S, Rady P, Orenstein J, Boto W, Kajumbula H, Romano N, *et al.* (2002). Genotypic analysis at multiple loci across Kaposi's sarcoma herpesvirus (KSHV) DNA molecules: clustering patterns, novel variants and chimerism. J Clin Virol *23*, 119–148

Zong J-C, Metroka C, Reitz MS, Nicholas J, and Hayward GS (1997). Strain variability among Kaposi sarcoma associated herpesvirus (HHV8) genomes: Evidence that a large cohort of U.S.A. AIDS patients may have been infected by a single common isolate. J Virol *71*, 2505–2511