Virus Evolution, 2019, 5(1): vey040

doi: 10.1093/ve/vey040
Research article

Evolutionary effects of the AID/APOBEC family of mutagenic enzymes on human gamma-herpesviruses

Teresa Martinez,[1] Maxwell Shapiro,[1] Sumita Bhaduri-McIntosh,[2,3] and Thomas MacCarthy[1,4,*]

[1]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794-3600, USA, [2]Division of Infectious Diseases, Department of Pediatrics, University of Florida, Gainesville, FL 32610, USA, [3]Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL 32610, USA and [4]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794-5252, USA

*Corresponding author: E-mail: thomas.maccarthy@stonybrook.edu

Abstract

The human gamma-herpesviruses, Epstein–Barr virus and Kaposi's sarcoma-associated herpesvirus, establish lifelong latency in B cells and are associated with multiple malignancies. Virus-host coevolution often drive changes in both host immunity and in the viral genome. We consider one host immune mechanism, the activation-induced deaminase (AID)/APOBEC family of cytidine deaminases, that induces mutations in viral DNA. AID, the ancestral gene in the family has a conserved role in somatic hypermutation, a key step in antibody affinity maturation. The APOBEC3 subfamily, of which there are seven genes in human, have evolved antiviral functions and have diversified in terms of their expression pattern, subcellular localization, and DNA mutation motifs (hotspots). In this study, we investigated how the human gamma-herpesviruses have evolved to avoid the action of the AID/APOBEC enzymes and determine if these enzymes are contributing to the ongoing evolution of the viruses. We used computational methods to evaluate observed versus expected frequency of AID/APOBEC hotspots in viral genomes and found that the viruses have evolved to limit the representation of AID and certain APOBEC3 motifs. At the same time, the remaining hotspots were highly likely to cause amino acid changes, suggesting prolonged evolutionary pressure of the enzymes on the viruses. To study current hypermutation, as opposed to historical mutation processes, we also analyzed putative mutations derived from alignments of published viral genomes and found again that AID and APOBEC3 appear to target the genome most frequently. New protein variants resulting from AID/APOBEC activity may have important consequences in health, including vaccine development (epitope evolution) and host immune evasion.

Key words: AID/APOBEC; gamma-herpesviruses; evolution.

1. Introduction

The human gamma-herpesviruses, Epstein–Barr virus (EBV) and Kaposi sarcoma-associated herpesvirus (KSHV), are large DNA viruses that establish chronic latent infections and have undergone prolonged coevolution with their human hosts (Cruz-Muñoz and Fuentes-Pananá 2017). Like all herpesviruses, the EBV and the KSHV life cycles have both lytic and latent stages. During latency, the viral genome replicates together with the host cell, protein expression is limited, and no virions are produced. The switch from latency to the lytic stage is essential for virus spreading and depends on the expression of key transcriptional activators and epigenetic modifications (Murata et al. 2012; Purushothaman, Uppal, and Verma 2015).

EBV is transmitted via saliva and infects ~95% of the human population by adulthood (Fields, Knipe, and Howley 2013). EBV infects epithelial and B cells, establishing latency primarily in B cells (Fields, Knipe, and Howley 2013). Although EBV infections are often asymptomatic, it has been implicated in several malignancies including Burkitt's lymphoma (particularly in Africa; Orem et al. 2007), nasopharyngeal carcinoma (particularly in Asian adults; Jemal, Bray, and Ferlay 1999), Hodgkin's lymphoma, gastric carcinoma, and breast cancer, among others (reviewed in Hippocrate, Oussaief, and Joab 2011; Young, Fah Yap, and Murray 2016). KSHV is transmitted via salivary or sexual routes (Fields, Knipe, and Howley 2013). This oncogenic virus has been associated with the development of Kaposi sarcoma, multicentric Castleman's disease and primary effusion lymphoma in immunocompromised patients. Transplant recipients are most susceptible to developing malignancies and non-neoplastic manifestation by KSHV primary infection or reactivation (reviewed in Riva et al. 2012). In contrast to EBV, KSHV can infect several cell types including B cells, epithelial cells, endothelial cells, monocytes and keratinocytes, and can establish latency in B cells and endothelial cells (Fields, Knipe, and Howley 2013).

Following infection, many innate immune mechanisms are induced including complement activation, apoptotic response and cytokine production. Both EBV and KSHV have evolved counteraction mechanisms. For example, both viruses express gene products that mimic cellular cytokines. In this study, we focus on one particular innate immunity mechanism—the Activation-induced deaminase (AID)/APOBEC family of DNA editing enzymes—because evaluating the effects on viral genome evolution is tractable.

AID/APOBEC enzymes have roles in both innate and adaptive immunity. Functionally, the enzymes deaminate deoxycytosine to deoxyuracil on ssDNA substrates (and in some cases RNA) which may be exposed during transcription or replication; the U:G mismatch then becomes a C to T transition mutation if replicated over in the absence of DNA repair (Morgan et al. 2004). The ancestral enzyme in the family, AID, is expressed in germinal center B cells and is required for the processes of somatic hypermutation and class-switch recombination that are part of antibody affinity maturation (Muramatsu et al. 2000; Petersen-Mahrt, Harris, and Neuberger 2002). AID deamination occurs preferentially at mutational 'hotspots' defined by the motif WRC (W = A/T, R = A/G) (Pham et al. 2003). The APOBEC3 (A3) subfamily enzymes act as part of innate immunity against exogenous viruses and endogenous retroelements (Harris and Dudley 2015; Willems and Gillet 2015). Primates have seven A3 genes (A–D, F–H), each of which has evolved specific subcellular localization and expression patterns (Conticello 2008). Furthermore, different A3 genes have evolved different mutational hotspots (Supplementary Table S1). For example, A3G preferentially mutates at CCC hotspots (Yu et al. 2004; Kohli et al. 2009), whereas for A3B it is TC (Bishop et al. 2004). However, a conclusive deamination preference has only been shown for AID and A3G. A recent study to determine the definitive hotspot motif for A3A and A3B was recently done using deep-sequencing, but this study failed to consider the possible biased effects of DNA repair enzymes (Shi et al. 2017). Another study showed a preferential motif of YYCR (Y =C/T, R = A/G) for A3A, but specific binding motifs depended on the pH level during the binding event (Pham et al. 2013). Furthermore, crystal structures of the A3 enzymes have been studied for hotspot preferences, but nucleotide positions 5′ and 3′ of the potential mutated cytosine were not definitively characterized (Salter and Smith 2018).

A role for A3 enzymes in viral restriction has been described for several retroviruses and DNA viruses (reviewed in Minkah et al. 2014; Harris and Dudley 2015; Willems and Gillet 2015). In the well-studied case of HIV, A3G targets nascent DNA during reverse transcription. However, HIV encodes the counteracting protein Vif, which targets the A3 enzymes for proteasomal degradation. Interestingly, several recent studies have suggested the possibility of a proviral role for A3G in which the enzyme also helps to generate escape variants (Wood et al. 2009; Monajemi et al. 2012, 2014). DNA viruses targeted by A3 enzymes include human papillomavirus, hepatitis B and C virus (HBV and HCV), herpes simplex virus-1 (HSV-1) and EBV (reviewed in Minkah et al. 2014; Harris and Dudley 2015; Willems and Gillet 2015). Overexpression of A3C has been reported to target the plus and minus DNA strands of HSV-1 and EBV lytic and latent genes, respectively (Suspene et al. 2011). Also, prolonged exposure to AID inhibits KSHV viral lytic reactivation and limits virion infectivity (Bekerman et al. 2013).

Several recent studies have shown that mis-regulation and mis-targeting of AID/APOBEC enzymes can contribute to the development of malignancies. In particular, mutation signatures associated with AID/APOBEC hotspots have been identified in Burkitt's lymphoma, breast and gastric cancer, among others (reviewed in Rebhandl et al. 2015). In EBV-infected cells, the EBV LMP-1 and EBNA-3C latent proteins upregulate AID expression by mimicking the active form of CD40 (Epeldegui et al. 2007), and by direct regulation of the AID (AICDA) gene, respectively (Kalchschmidt et al. 2016). KSHV infected B cells also show increased AID expression (Bekerman et al. 2013). Gene expression data from the GTEx Consortium show an increase not only in the mRNA levels of AID but also in almost all of the A3 enzymes (except for A3A) in EBV-transformed lymphocytes when compared with the mRNA levels observed in spleen cells (see Supplementary Fig. S1).

Given the existing evidence associating certain mutations in the virus with tumor development, it is important to identify the mechanisms driving genetic diversity and evolution in human gamma-herpesviruses (Rebhandl et al. 2015). Previous work by ourselves used computational methods to address the question of why these enzymes evolved their particular mutational hotspots (Chen and MacCarthy 2017). Here, we consider in more detail the impact of the AID/A3 family on the evolution of the human gamma-herpesviruses. Human gamma-herpesviruses have undergone a prolonged coexistence with humans and would be expected to have evolved as much as possible to limit the detrimental effects of AID/APOBECs, or potentially use them for their benefit by generating adaptive variation. Assuming that a prolonged coevolution has already shaped the evolution of the viruses, then experiments involving AID/A3 gene knockouts will most likely be uninformative because AID/A3 effects are expected to be small. We consider that analyzing viral genomes for the evidence of previous evolution in the presence of AID/A3 will be more informative. In this study, we used computational methods to investigate the extent to which the EBV and KSHV genomes (1) have evolved to limit the number of hotspot motifs to avoid AID/A3 mutations, and (2) to determine the hypermutation due to AID/A3 in these viruses.

2. Materials and Methods

2.1 Dataset

For this analysis, we used complete or near complete viral genome sequences without ambiguous nucleotides. The DNA

sequence of the EBV and KSHV viral genomes analyzed were obtained from NCBI-Nucleotide Database. Supplementary Table S2 lists the accession number, strain name, and country of isolation.

2.2 Quantification of hotspot representation

The quantification of AID and A3 hotspot representation in the gamma-herpesviruses was determined using the cytidine deaminase representation reporter (CDUR) (Shapiro, Meier, and MacCarthy 2018). This reporter consists of different shuffling algorithms that generate a null distribution of 1,000 random sequences from a subject sequence, and compares the number of hotspots in the sequence of interest to this null model. The null model is generated differently for coding versus noncoding regions. For noncoding sequences, we used the random module in Biopython to generate 1,000 shuffled sequences that strictly conserved the GC content of the DNA sequence as variations in GC content itself can lead to different representations of A3 hotspots (Chen and MacCarthy 2017). For coding sequences, we used the n3 sequence shuffling; this method shuffles the third position nucleotide in the codon while maintaining the encoded protein sequence. Because this method shuffles the nucleotides in the sequence of interest, it maintains the overall GC content of the coding sequences. However, it changes the codon usage bias. The *P*-value for hotspot under-representation was calculated as the proportion of null model sequences with fewer than *n* hotspots, where *n* was the number of hotspots in the actual sequence. Over-representation was measured as 1 minus the reported *P*-value for under-representation. The set of *P*-values computed for the coding sequences were adjusted for multiple comparisons using the Benjamini–Hochberg False Discovery Rate (FDR) test. Sequences with a *P*-value $\leq$0.05 were considered as under-represented while those with a 1 $-$*P*-value $\leq$0.05 (i.e. a *P*-value $\geq$0.95) were considered as over-represented for a particular hotspot. The hotspots analyzed were WRC (W = A/T and R = G/A) for AID (Pham et al. 2003a,b); and CCC for APOBEC-3G (A3G) (Yu et al. 2004; Holden et al. 2008; Kohli et al. 2009); as well as TC and TTC for the remaining APOBEC3s and APOBEC1 (Bishop et al. 2004; Kohli et al. 2009; Hultquist et al. 2011; Suspène et al. 2013; Saraconi et al. 2014; Adolph et al. 2017).

2.3 Hypermutation analysis

To determine the significance of the mutations and the ongoing evolution due to AID and A3 enzymes we used hyperfreq, a method described by Matsen et al. (2014). This method uses a Bayesian two-context mutation probability ratio to determine if the number of mutations in a given context (AID/A3 hotspots) is higher than out of context mutations. We ran hyperfreq analyses in four different contexts: GGG, GAA, GA and GYW, the reverse complements of the APOBEC3, APOBEC1, and AID hotspots (Supplementary Table S1). To evaluate if there is a mutation strand bias in the human gamma-herpesviruses, we then ran hyperfreq on both DNA strands (template and coding strand for the protein coding sequences; top and bottom strands in noncoding sequences). Statistical significance was determined using the one-sided Fisher exact test. The protein coding sequences were collected directly from NCBI and aligned using the TranslatorX server and ClustalW as the alignment tool. TranslatorX translates the nucleotide CDS into amino acids, aligns them, and then back-translates them to DNA nucleotides (Abascal, Zardoya, and Telford 2010). Noncoding sequences were aligned using the command-line version of ClustalW. In each case, the resulting sequences were used as input to hyperfreq. We used reference sequences NC_007605 and NC_009333 for EBV and KSHV, respectively, for the comparison in hyperfreq (see Supplementary Fig. S6 for an example of a hyperfreq alignment).

3. Results

3.1 Gamma-herpesviruses have evolved to limit WRC and TC motifs

As noted in the Introduction, AID and other A3 genes are upregulated as part of the Interferon response (Bonvin et al. 2006), as well as in EBV-transformed B cells. As suggested in a previous study (Chen and MacCarthy 2017), continuous evolution with exposure to AID/APOBEC enzymes may drive a reduction in the number of enzyme hotspots in the viral genomes in order to reduce detrimental mutations, particularly in coding regions. This previous study introduced a method to quantify whether there was a statistical reduction (under-representation) of AID/A3 family motifs in viral genomes. Using an updated version of this method, we analyzed a given coding sequence by comparing the number of hotspots (e.g. CCC hotspots for A3G) with a null distribution based on sequences that preserve the amino acid content but shuffle the codons used. If the original sequence falls on the left tail of this distribution, it indicates hotspot under-representation, whereas sequences on the right tail indicate potential over-representation. Because previous results suggested that GC content was a major factor in determining the hotspot under-representation, we used a shuffling method that exchanges nucleotides within a sequence at the third codon position only and therefore does not change GC content (see Section 2). The method is available as a separate software package (CDUR) (Shapiro, Meier, and MacCarthy 2018).

We used CDUR to quantify AID and A3 hotspot under-representation for the EBV and KSHV reference genomes (NC_007605 and NC_009333). We also hypothesized that any effects on viral evolution might act at different stages of the viral life cycle, and therefore considered distinct lytic life cycle stages, as well as latency separately (Lu et al. 2006). Figure 1 shows the fraction of EBV and KSHV coding sequences in each category with under- or over-representation of TC and TTC hotspots. We found a strong signal for under-representation for TC motifs in both EBV and KSHV. To verify that TC hotspot under-representation was specific to the TC motif, we compared the TC under-representation profile with the other three NC alternative (AC, GC, or CC) profiles and indeed found that under-representation was far greater for TC than for these other motifs (Supplementary Fig. S2). There was no observable trend within the lytic cycle stage for either EBV or KSHV, and both viruses had high under-representation for latency-associated genes, in particular KSHV where five of its six latency genes (83%) were under-represented for TC hotspots. Both EBV and KSHV infect and undergo their lytic cycle in epithelial cells, where A3B and A3H, which both recognize TC motifs, are expressed at high levels (A3A is most likely not implicated due to having low expression levels). The lytic cycle involves high rates of transcription, at least of lytic genes, and replication both of which may increase the probability of exposing ssDNA to A3 enzymes. The pattern of under-representation for TTC motifs (Fig. 1B) was very similar to that of TC motifs in EBV (Fig. 1A). This may result, at least in part, as side effect of strong under-representation for the TC motif, since the TC motif is embedded

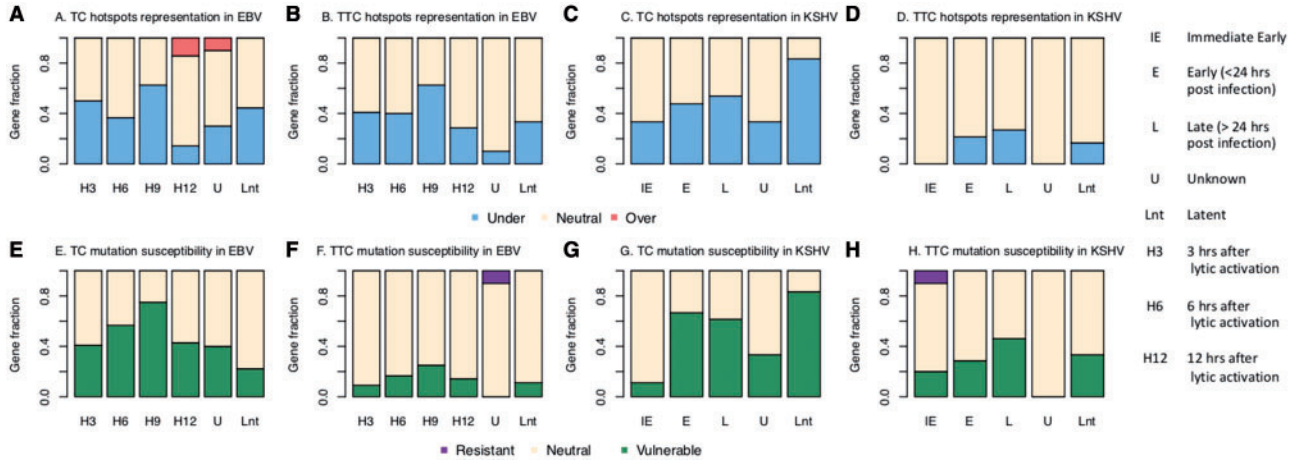

Figure 1. TC and TTC hotspot representation and vulnerability in human gamma-herpesvirus genes separated according to their transcription profile kinetics. Upper panel: hotspot representation; lower panel: mutation vulnerability. Under-represented or resistant genes are those with a *P*-value <0.05; while those with a 1−*P*-value ≤0.05 (i.e. a *P*-value ≥0.95) were considered as overrepresented or vulnerable. *P*-values obtained were adjusted using the Benjamini–Hochberg method. Resistant genes are those in which the action of the cytidine deaminases will have a higher incidence of synonymous changes, susceptible genes those with a higher incidence of non-synonymous changes, and neutral genes are those with an expected number of nonsynonymous mutations. Genes with under-representation and vulnerable genes are not necessarily the same (see Supplementary Fig. S4 for genes that overlap). EBV and KSHV transcription profiles as determined by Lu et al. (2006) and reviewed in Dourmishev et al. (2003), respectively.

within TTC, leading to a positive correlation between the two motifs (Table 1). However, the correlation is modest and the similarity between TC and TTC is not evident for KSHV, which suggests some direct effect of under-representation for TTC motifs in the case of EBV.

In addition to hotspot frequency, another factor to consider is the potential impact of hotspot mutations, in particular whether the mutations cause an amino acid change (nonsynonymous) or not (synonymous mutation) (Chen and MacCarthy 2017). The CDUR package evaluates the ratio of non-synonymous mutations to the number of observed hotspots, again comparing the observed ratio with that of the null distribution using the n3 codon shuffling model for each coding sequence (the statistic is labeled as 'repTrFrac', short for 'replacement transition fraction' in CDUR; see Section 2). Here, for a given gene, if the observed ratio falls on the left tail of the distribution, that is, fewer than expected nonsynonymous mutations, then we describe the gene as being mutationally 'resistant', whereas if it is on the right tail, we describe the gene as mutationally 'vulnerable' (Chen and MacCarthy 2017). We hypothesized that if a viral genome was under A3 mutational pressure during its evolution, not only would the number of hotspots be reduced, but also the remaining hotspots would predominantly cause nonsynonymous replacements if mutated, making the gene mutationally 'vulnerable'. In other words, by reducing the number of hotspots that cause synonymous mutations, the proportion of hotspots which may deleteriously alter the protein would increase. For example, it has been shown in HIV-1 sequences that patients with a nucleotide sequence of TGGG, that is, a tryptophan codon (TGG) followed by a G, is susceptible to A3G editing. In such a case, the resulting mutation (G > A in the underlined G) leads to a premature stop codon (TAG), rendering the resulting protein nonfunctional (Cuevas et al. 2015). Furthermore, it was shown in this study that patients having this mutation had a better clinical outcome. In agreement with our hypothesis, we found that both EBV and KSHV are strongly enriched in significantly 'vulnerable' genes for the TC hotspot (Fig. 1E and G). Again, the profile for the TTC hotspot is similar to TC, but attenuated

Table 1. TC motifs are negative correlated with WRC and CCC motifs. Mean and standard deviation for the correlation of hotspots representation in EBV and KSHV genes.

Hotspot	CCC	WRC	TC
EBV genes			
TTC	0.037 ± 0.040	−0.102 ± 0.064	0.179 ± 0.119
CCC		−0.065 ± 0.065	−0.181 ± 0.056
WRC			−0.383 ± 0.056
KSHV genes			
TTC	−0.000 ± 0.039	−0.124 ± 0.049	0.267 ± 0.061
CCC		−0.068 ± 0.054	−0.200 ± 0.051
WRC			−0.395 ± 0.048

(Supplementary Fig. S4 shows the number of vulnerable genes that also have an under-representation of TC or TTC hotpots).

Although both EBV and KSHV are latent in B cells and these are the primary human cell type to express AID at high levels (albeit only at the germinal center stage), we observed only limited under-representation of AID WRC hotspots in latency genes in KSHV and none at all in EBV (Supplementary Fig. S3A and C). Interestingly, there is a relatively strong negative correlation between WRC and TC motifs (Table 1), which is a consequence of the contexts at the −1 position (R = A/G vs T) being mutually exclusive. Thus, there is a tradeoff between APOBEC3 TC hotspots and AID WRC hotspots in which APOBEC3 may dominate if, for example, APOBEC3 acted strongly on all genes, including latency genes, during lytic replication.

Specifically, in EBV, there is an over-representation of CCC motifs (A3G), particularly for late lytic genes (H9, Supplementary Fig. S3B). Hotspot over-representation is less straightforward to interpret in terms of its mechanistic and evolutionary consequences. It may arise to promote diversity, as has been proposed in the context of HIV (Sadler et al. 2010), or it may also arise as a consequence of correlations with other motifs. CCC motifs are negatively correlated with TC motifs,

and to an extent with WRC (AID) (Table 1). Even while there is an increase in A3G mRNA levels in EBV-transformed lymphocytes (Supplementary Fig. S1B) and a higher probability of being mutated by enzyme due to the over-representation, mutations due to A3G may not be as detrimental for the virus given that A3G has been shown to be predominantly cytoplasmic which suggests that the virus, which replicates in the nucleus, may not be as exposed to A3G as it would be to AID or other APOBEC3s. Furthermore, since the virus may need a certain GC content for its functions, it would be advantageous in this case to place a large portion of its G/C nucleotides in CCC motifs which would less likely be mutated, though this has not been shown. EBV and KSHV have a CDS mean GC content of 59 and 53% in coding regions, which may suggest a functional need for a higher GC content. One example for EBV is the nucleotide sequence in the Gly–Ala repeat region of the EBNA-1 protein, which has a strong codon bias (using GCA for Ala and GGR for Gly). Furthermore, this repeat region is critical to inhibit antigen presentation through the major histocompatibility complex (MHC) class 1 pathway (Levitskaya et al. 1995; Apcher et al. 2010). In our analysis, we found that mutations caused by A3G would result in non-synonymous amino acid changes in this region of EBV EBNA-1 latency-associated protein that would not be tolerated physiologically. In other sequences, there may be a need for specific mRNA secondary structures that may drive a particular GC content, as it has been shown for RNA viruses (Simmonds, Tuplin, and Evans 2004). Almost all KSHV genes have a neutral representation of A3G hotspots, consistent with the fact that the type of cells KSHV infects is much broader than EBV.

Genes that have a depletion of TC motifs include those involved in virus DNA replication, host immune system regulation, control of lytic gene expression, and virion function and structure (Supplementary Table S3). The genes overlap to a large extent with those depleted of TTC motifs. Notably, the latency-associated proteins EBNA-2 and -3A showed a depletion of TC and TTC hotpots; however, of the LMP genes, only LMP-2B showed depletion for TC hotspots. To determine if the different LMP1 variants described in EBV associated malignancies differ in AID/A3 hotspot representation, we analyzed the six different LMP1 variants described by Edwards et al. (2004) for hotspot representation. No difference in TC, TTC, or WRC hotspot representation was observed when compared with the reference strain NC_007605; however, four variants (NC, Alaskan, China 1, and China 2) have a depletion of CCC hotspot.

Two key immediate-early genes of EBV, BZLF1, and BRLF1 (that code for the transcriptional activators Zta and Rta, respectively) were also significantly depleted for TC hotspots. These transcriptional activators are sufficient to induce lytic replication in EBV latently infected B cells or epithelial cells. In KSHV, along with genes with significant depletion of TC motifs also overlap with EBV in terms of function; in particular, the majority of KSHV latency-associated proteins, including LANA-1 (Supplementary Table S4). While this article was under review, a new study was published showing that the EBV lytic gene BORF2, can bind to and stoichiometrically inhibit the deaminase activity of A3B (Cheng et al. 2018). Interestingly, we found that BORF2 was significantly underrepresented for both TC and TTC motifs ($P < 10^{-3}$ and $P = 0.014$, respectively, FDR corrected), which would be consistent with minimizing the potential for APOBEC mutations occurring before the availability of BORF2 protein. ORF 61, the BORF2 ortholog in KSHV, was also found to have underrepresentation for both these motifs ($P = 0.015$ and $P = 0.029$, respectively). Both EBV BORF2 and KSHV ORF 61 were also classified as significantly susceptible with respect to amino acid changes, lending further support to their having evolved under mutational pressure. Overall, our results suggest that the human gamma-herpesviruses have evolved to reduce the presence and impact of TC hotspots and, to a lesser extent, for TTC hotspots. The over-representation of CCC (A3G) hotspots may in part arise from a negative correlation with the relatively strong signal for TC hotspot under-representation.

Table 2. Hypermutated genes. Total count of genes shows genes that were hypermutated in at least one of the EBV or KSHV genomes used in this study.

	Hypermutated genes	Not hypermutated genes
EBV		
TC	32	50
TTC	27	55
WRC	31	51
CCC	22	60
KSHV		
TC	33	49
TTC	26	56
WRC	35	47
CCC	21	61

3.2 WRC and TC motifs in the CDS in EBV are targeted more frequently than TTC or CCC motifs

In the analysis above we considered the profile of A3 hotspots, which is informative as to the evolutionary history of these viruses but does not provide evidence for any hypermutation trends, if any exist. A reduction in hotspots for one APOBEC enzyme, A3B, for example, might increase the relative probability for another, such as AID. In a recent *ex vivo* study, Suspene et al. (2011) showed A3C hypermutation of the latency associated EBNA-1 and -2 genes in human PBMC lines and *in vivo* from a patient, respectively. However, the impact of A3C and other cytidine deaminases in other EBV coding sequences, as well as in KSHV remains unknown. We employed a Bayesian statistic method (hyperfreq) as described by Matsen et al. (2014) to look for AID/A3-dependent hypermutation in EBV and KSHV coding sequences (see Section 2). Table 2 summarizes the number of genes that showed the evidence of significant AID/A3 hypermutation, according to the hyperfreq package, in at least one of the nine EBV genomes used in this study. We identified a slightly higher number of TC-dependent and WRC (AID)-dependent hypermutated genes than TTC and CCC (A3G), although not statistically significant. Given that CCC-dependent mutation appears to be less common, this again suggests that although the virus is over-represented for CCC motifs, it is not frequently exposed to the enzyme.

Next, we analyzed the specific genes that were targeted by AID/A3 enzymes. Table 3 lists the EBV genes that showed AID- or A3-dependent hypermutation in 40% or more of the genomes analyzed, their function and transcription profile during the lytic cycle (see Supplementary Table S5 for the complete list of genes that showed evidence of hypermutation). Genes with AID-dependent (WRC) hypermutation included genes involved in the production of new virions (BFRF1, BFRF1A, BFRF2, BFLFL2, and BDLF4). BFRF2 and BDLF4 (together with four other proteins) form a complex required for the expression of late genes. BFLFL2 and BFRF1 form the nuclear egress complex that is required for the exit of the assembled capsid and BFRF1A is

Table 3. AID- and APOBEC3-dependent hypermutation in EBV genes.

Gene	Transcription profile[a]	Function	Motif (hypermutated genomes)
BFRF2	3 h	Mediate late gene transcription (Djavadian, Chiu, and Johannsen 2016)	WRC (6)
BDLF3	6 h	Enhances epithelial infection, virion protein (Fields, Knipe, and Howley 2013)	TC (6)
BNRF1	Unknown	Major tegument protein (Fields, Knipe, and Howley 2013)	TC (5)
LF1	12 h	Uncharacterized protein	TC (5)
BRRF2	6 h	Production of infectious progeny (Watanabe et al. 2015)	TTC (5)
BDLF4	6 h	Required for expression of late genes (Fields, Knipe, and Howley 2013)	WRC (5)
BALF4	9 h	gB-fusion protein, virion protein (Fields, Knipe, and Howley 2013)	TTC (5), WRC (4)
LMP1	Latent		TC (4), TTC (4)
BFLF2	6 h	Required for Nuclear Egress—Binds BFRF1 (Fields, Knipe, and Howley 2013)	TTC (4), WRC (4)
BALF2	3 h	Single stranded DNA-binding protein (Fields, Knipe, and Howley 2013)	WRC (4), CCC (4)
EBNA3B	Latent		CCC (4)
BFRF1	3 h	Essential for primary viral envelopment and egress (Farina et al. 2005)	WRC (4)
BFRF1A	Unknown	DNA packaging (Pavlova et al. 2013)	WRC (4)
BGLF1	3 h	Capsid maturation/DNA packaging, Virion protein (Fields, Knipe, and Howley 2013)	WRC (4)
LF2	6 h	Immune evasion and regulation of lytic activation (Wu et al. 2009; Heilmann, Calderwood, and Johannsen 2010)	TTC (4)

[a]Hours after lytic induction as determined by Lu et al. (2006).

involve in DNA packaging. Targeting of these genes by AID is consistent with their expected expression in B cells. Certain genes (BALF2, BALF4 and BNRF1) showed both TC-dependent hypermutation and TC hotspot under-representation (Table 2 and Supplementary Table S3) suggesting both past and ongoing mutational targeting. The BALF2 gene, which also showed TTC-dependent hypermutation, codes for a ssDNA-binding protein, and therefore mutations in this gene could affect their DNA binding leading to longer exposure of the ssDNA to mutagenic enzymes such as the AID and A3 enzymes. Given the importance of the BALF2 gene, we compare the number of mutated CDSs in genomes with and without hypermutated BALF2 gene. We did not find a correlation between mutated BALF2 genes and higher number of mutated CDSs; however, the BALF2 gene showed both TC hotspot under-representation and vulnerability to TC hotspot mutations suggesting that exposure to A3 could indeed affect their DNA-binding affinity. The BRRF2 gene, which is important for the production of virions, showed TTC-dependent hypermutation. TTC-dependent hypermutation was also observed in the BALF4 gene, which codes for glycoprotein B, and mutations in this gene could affect residues that may have important roles in virus specific entry. Together these results suggest that the AID/A3 enzymes are targeting some EBV lytic genes that are important for the production of new virions.

The APOBEC enzymes preferentially target ssDNA cytosines on the coding strand of the transcription bubble or those on the lagging strand of the replication fork (Bhagwat et al. 2016; Hoopes et al. 2016). To determine if the AID/A3 enzymes target the coding sequences during transcription, then we consider the evidence of hypermutation in both template and coding strands (Supplementary Table S6). We identified a significant preference of AID-dependent hypermutation for the coding strand (*P*-value 0.01656) in EBV genomes, as would be predicted if AID was acting during transcription. Also, there seems to be a slight preference for hypermutation in the coding strand by A3G, although the difference is not statistically significant. We performed this analysis, but did not find any statistically significant bias for the other APOBECs.

3.3 AID and A3 enzymes target mostly early and late expressed genes in KSHV

We next considered KSHV and used the CDS of 18 KSHV genomes available in Genbank to look for the evidence of AID/A3 enzyme hypermutation. Table 2 shows the number of genes that were hypermutated in at least one of the 18 KSHV genomes. Similar to EBV, the number of AID- and TC-dependent hypermutated genes was slightly higher than those with TTC and CCC hotspot mutations, however, they were statistically significant only for CCC (TC vs CCC *P*-value 0.01592 and WRC vs CCC *P*-value 0.03354). Table 4 shows the hypermutated genes in 40% or more of the genomes, their function and transcription profile (see Supplementary Table S7 for complete list). Hypermutation was mostly observed in early or late genes (Table 4). Genes with WRC (AID) hypermutation were mostly associated with host immune regulation. ORF75, which also plays a role in immune evasion, was hypermutated at TC and TTC hotspots and also showed under-representation for TC and TTC. Except for AID, the number of hypermutated sequences in the template strand was slightly higher than the coding strand; however, the difference was not statistically significant in any of the cases (Supplementary Table S8). Altogether these results suggest that AID/A3 enzymes are predominantly targeting genes involved in host immune evasion.

3.4 AID targets the oriP origin of replication sequence in EBV

During latency, the origin of replication (oriP) site is responsible for the initiation of replication and maintenance of the episome. EBV contains a single oriP site, which consists of two functional domains that contain EBNA-1-binding sites: the dyad symmetry sequence and the family of repeats (FR). The interaction between the major latency protein (EBNA-1) and the oriP sequence is essential for the stability and replication of the viral episome (reviewed in Hammerschmidt and Sugden 2013). Given its importance, we sought to evaluate AID/A3 hotspot under-representation and evidence of hypermutation for this particular noncoding subregion. Because oriP is a noncoding sequence

Table 4. AID- and APOBEC3-dependent hypermutation in KSHV genes.

Gene	Transcription profile[a]	Function	Motif (hypermutated genomes)
K14	Early	Immune evasion (reviewed in Purushothaman, Uppal, and Verma 2015)	WRC (11)
ORF63	Early	Immune evasion (reviewed in Purushothaman, Uppal, and Verma 2015)	WRC (8)
ORF22	Late	Glycoprotein H (Fields, Knipe, and Howley 2013)	WRC (13)
ORF54	Early	Immune evasion (Madrid and Ganem 2012)	WRC (8)
ORF49	Early	Transcription factor (Fields, Knipe, and Howley 2013)	TTC (8)
ORF75[b]	Late	Immune evasion (Purushothaman, Uppal, and Verma 2015)	TTC (10), TC (9)
ORF68	Late	Virion glycoprotein (Fields, Knipe, and Howley 2013)	CCC (8)

[a]Transcription profile as reviewed in Dourmishev et al. (2003). Early, <24 h postinfection; late, >24 h postinfection.
[b]Seventeen genomes.

we used a random shuffle to evaluate the under-representation (see Section 2). Largely consistent with the trends we observed in coding regions, we found that the oriP sequence has an under-representation for TC and TTC motifs, a neutral representation of WRC motifs and over-representation of CCC motifs (Supplementary Table S9).

We next evaluated if there was evidence of hypermutation in the EBV oriP sequence. The number of EBNA-1-binding sites, as defined by the motif GRWWRYVYRYVCTDYY (Dresang, Vereide, and Sugden 2009), in the genomes analyzed varies from 19 to 28, most of which belong to the FR repeats region. To avoid possible false positive results arising from hotspots within the FR repeats, we removed the FR repeats from the sequences before generating the sequence alignment. We found the evidence of TC-dependent hypermutation on both strands of the oriP sequences in four of the EBV genomes. We also identified two genomes with WRC- and TTC-dependent hypermutation in the top and bottom strands, respectively (Supplementary Table S10), but no evidence of CCC hypermutation. Figure 2 shows the mutation frequency (frequency of minor alleles) that arises from the alignment. The bottom part of the figure also shows the positions of the AID/A3 hotspots based on the alignment consensus sequence. As shown in the figure, there is a high density of AID hotspots (WRC and its reverse complement motif GYW) throughout the entire EBV oriP sequence. However, the FR (nucleotides 138–715) has a far higher mutation frequency than the rest of the sequence. Interestingly, given the EBNA-1-binding site consensus sequence (GRWWRYVYRYVCTDYY), AID-dependent mutations at WRC hotspots within the FR region appear to strongly avoid affecting the integrity of the EBNA-1-binding sites. Within this relatively short sequence there are twenty-six possible WRC hotspots in the EBNA-1-binding site consensus sequence (taking into account the ambiguous nucleotides and considering both strands) and of these, twenty-five can tolerate a C–T mutation. Unfortunately, for KSHV there is only one oriP sequence in Genbank so we were unable to assess AID/A3 hypermutation.

During the lytic cycle, EBV replication starts from the oriLyt. EBV usually have two oriLyt copies that contain Zta response elements to which the transcriptional activator Zta bind and recruit other viral replication proteins (Schepers, Pich, and Hammerschmidt 1993; El-Guindy, Heston, and Miller 2010). Given the importance of this region for the viral replication and spreading, we evaluate the AID/A3 hotspot representation and hypermutation. Similar to the trends observed in the rest of the viral genome we found an under-representation of TC and WRC hotspots, neutral representation of TTC hotspots, and an under-representation of CCC hotspots (Supplementary Table S11 and Fig. S5). Once again, the over-representation of CCC hotspots could be as a result of the avoidance of other A3 and AID enzymes and the fact that the virus is not exposed as often to A3G enzymes. We did not find evidence of hypermutation due to AID/A3 enzymes in the oriLyt sequenced for any of the analyzed genomes. These results, once again, suggest that EBV has evolved to limit the detrimental effects of both AID and TC-targeting A3 enzymes.

4. Discussion

We have analyzed human gamma-herpesvirus genomes to investigate the evolutionary consequences of exposure to AID/A3 family enzymes. Our results suggest that human gamma-herpesviruses, in general, are under evolutionary pressure, and are subject to AID/A3 hypermutation. Moreover, human gamma-herpesviruses have evolved to limit the action of AID/A3 enzymes. In a previous study, we found that hotspot under-representation measures could be affected by several factors apart from direct hotspot avoidance, including codon bias and especially GC content (Chen and MacCarthy 2017). For this study, we therefore adopted a new method (CDUR; Shapiro, Meier, and MacCarthy 2018) in which the shuffling method used to generate the null model strictly maintains the GC content while introducing minimal changes to the codon usage bias.

In response to an infection, host cells may coexpress several AID/A3 enzymes (together with activation of many other immune pathways, of course). As a consequence of differences in the hotspots of the AID/A3 enzymes, many of which are mutually exclusive, it would be extremely difficult for any virus to evolve under-representation for all hotspots simultaneously. When analyzing genomes for hotspot under-representation we also considered whether those mutations would be synonymous or nonsynonymous, under the hypothesis that AID/A3 evolutionary pressure would both reduce hotspot number and increase the impact of the remaining 'indispensable' hotspots. This dual signal was observed most strongly for TC motifs. Many EBV and KSHV genes are thus vulnerable to A3-mediated amino acid changes at TC motifs, including the latency associated EBNA-3A and -3B in EBV and the majority of KSHV latency proteins. Specifically, for the latency proteins, the amino acid changes may lead not only to protein dysfunction, but also to cancer. For example, because LMP-1 activity is critical for latency maintenance and induces NF-kB activity, it has been suggested that its disruption or single amino acid variations may drive tumors (Bentz, Shackelford, and Pagano 2012; Renzette et al. 2014).

Under-representation for one AID/A3 motif may lead to under- (or over-)representation of another motif as a consequence of correlations, both positive and negative, between different motifs. Unsurprisingly, the TC motif is positively correlated with the motif TTC, which contains a TC. Both TC and TTC are

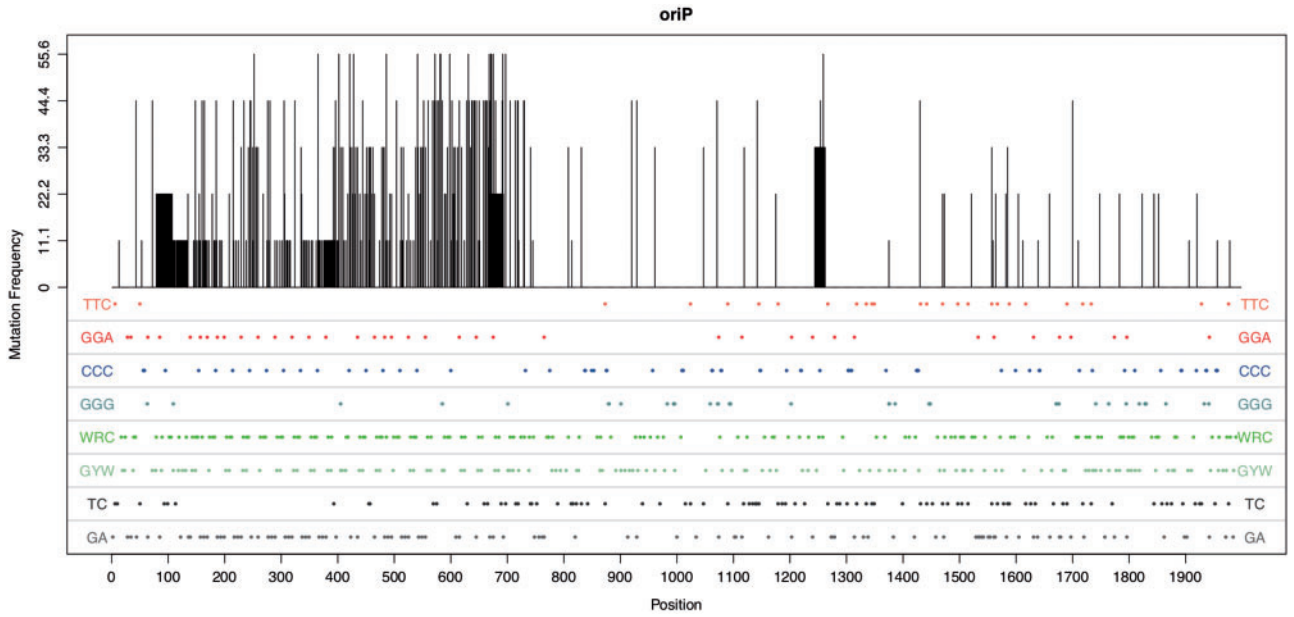

Figure 2. Mutation frequency and distribution of the AID and APOBEC3 activity in the EBV oriP sequence. The x-axis shows the hotspots sites within the gene. The y-axis represents the mutation frequency for each site. Each colored dot in the bottom panels represents an AID or A3 hotspot (or their reverse complement sequence) as labeled at both edges.

negatively correlated with the nonoverlapping motif CCC. Thus, the over-representation we observe for the CCC motif (A3G) in both EBV and KSHV may well be a consequence of TC under-representation rather than direct selection. Our results suggest that A3B may be the primary deaminating A3 enzyme due to the predominant nuclear localization of A3B (Salter, Bennett, and Smith 2016) compared with the other APOBEC3s, as well as high relative expression (A3B is expressed >6-fold higher in EBV-transformed lymphocytes versus spleen cells—Supplementary Fig. S1B). However, we cannot rule out the fact that other APOBECs are responsible which have also been shown to have nuclear localization (Muckenfuss et al. 2006; Lackey et al. 2012; Salter, Bennett, and Smith 2016) may be responsible. Even though A3B and A3H enzymes recognize the same hotspot (TC) as do A3C and A3F enzymes (TTC), it would appear to be more likely that A3B and A3C enzymes are the ones targeting the gamma-herpesviruses genomes for biological reasons. However, as stated previously, we cannot completely rely on hotspot motifs to conclusively differentiate between APOBEC3s, save for A3G. A3B is predominantly nuclear, while A3H shuttles between the nucleus and the cytoplasm; also, A3H tends to oligomerize in the cell (Lackey et al. 2013; Gu et al. 2016; Ito et al. 2017). A3C can be found in the nucleus and cytoplasm while A3F has been shown to be strictly cytoplasmic (Lackey et al. 2013). Many studies have shown that A3B is constitutively nuclear (Bogerd et al. 2006; Stenglein and Harris 2006; Stenglein, Matsuo, and Harris 2008; Pak et al. 2011; Lackey et al. 2012, 2013; Salter and Smith 2018), and this nuclear import is determined by distinct sets of residues in the N-terminal domain of the enzyme (Salamango et al. 2018). A3H has been shown to be primarily cytoplasmic, and that RNA binding is required for this cytoplasmic localization (Shaban et al. 2018). However, this was shown by using RNA-binding mutants in the tumor cell lines 293T and HeLa cells. Other studies have also shown that A3H is primarily cytoplasmic (Hultquist et al. 2011; Li and Emerman 2011), but these and other (Starrett et al. 2016) studies also showed the existence of A3H haplotypes that have nuclear localization. A3C has been shown to shuttle between the nucleus and cytoplasm, while A3F has been shown to be predominantly cytoplasmic during both interphase and mitosis (Lackey et al. 2013). The results of these studies, together with the median expression levels reported (Supplementary Fig. S1), suggest it may be more likely that A3B and A3C are the enzymes responsible for deaminating gamma-herpesviruses, as those have both been shown to have greater nuclear localization than A3H and A3F, respectively. However, we cannot rule out the possibility of A3H and A3F enzymatic activities as the sources for the reported mutations. At the same time, most of the studies cited above used overexpression of the respective APOBEC3 enzyme, some in nonrelevant cell types, to show nuclear or cytoplasmic localization. As such, further exploration is required to confirm that A3B and A3C are indeed the enzymes responsible for targeting gamma-herpesviruses.

The CCC motif over-representation could also be a decoy mechanism that the EBV has evolved to sequester the AID enzyme. Biochemical studies have established that AID avoids 'SYC' motifs (S = G or C and Y = T or C) (Peled et al. 2008; Chelico, Pham, and Goodman 2009). The CCC motif is embedded within these 'coldspots', and these instances of over-representation of CCC, which is accompanied with high resistance, could be used as a decoy mechanism to sequester the AID enzyme and limit the low frequency of mutations to specific regions of the viral genome. Something similar might be happening with the high fraction of genes that showed an over-representation of CCC motifs.

Even though both EBV and KSHV increase the level of AID expression in infected B cells we did not observe an under-representation of AID hotspots. Both viruses may have evolved various mechanisms to reduce AID action directly. When expressed, the EBV latency protein EBNA-2 can inhibit AID expression (Tobollik et al. 2006); EBV also causes increased expression of endogenous miR-155 and miR-93 (Zhu et al. 2014; Namba-Fukuyo et al. 2016), which prevent AID translation (Dorsett et al. 2008; Linnstaedt et al. 2010; Borchert, Holton, and

Larson 2011). KSHV uses LANA-1 to recruit UNG2 (Uracil DNA glycosylase 2) which may neutralize the mutagenic effects of AID; also, the virus uses two miRNAs (K12-11 and K12-5) to again repress AID translation (Vieira et al. 2013).

Even though both EBV and KSHV have mechanisms to reduce AID-dependent mutations as we have discussed, our results suggest that AID, together with certain A3 enzymes, are in fact editing the viral genomes. Since the substrate for these enzymes is ssDNA, it has been suggested that there will be more lagging or coding strand mutations than in the leading or template strand during replication and transcription, respectively (Suspene et al. 2011; Hoopes et al. 2016). However, we found that AID and A3 enzymes target both strands in human gamma-herpesviruses in accordance with previous observations by Suspene et al. (2011). This result could be explained by the fact that EBV episome replication can start from multiple sites, not only from the oriP, as demonstrated by Norio and Schildkraut (2004). Moreover, in this last study, the authors not only observed changes in the duplication speed and replication fork movement in the DNA segments replicated, but also that the frequency of initial replication from the oriP varies among strains. In light of these previous results, it is not possible to confirm which CDSs are within the lagging or the leading strand during episome replication. Another fact that may explain why we observed hypermutation in the template and coding strands could involve the unidirectional replication of EBV when the replication starts from the oriP. It is known that EBNA-1 causes a stalling of the replication fork in the FR region within the oriP sequence (Dhart and Schildkraut 1991), and when this happens the replication fork can move such a way that the top is always the lagging strand and the bottom strand is leading. As a result, some of the sequences that serve as the template strand during transcription will be in the lagging strand during replication, and vice versa. If the mutations arise when the episome is replicating only from the oriP the AID/A3 enzymes may be competing with the virus ssDNA-binding proteins as suggested by Willems and Gillet (2015). The same explanation may apply to KSHV since we saw mutations in both strands and it was recently shown that LANA-1 protein also causes a stalling of the replication fork at the terminal repeat regions (Dheekollu et al. 2013). During canonical somatic hypermutation in B cells, AID activity is associated with transcription rather than replication of the DNA (Larson and Maizels 2004); however, in a recent study, it was shown that AID mutagenic activity is restricted to the early G1 phase of the cell cycle (Wang et al. 2017). The fact that EBV lytic reactivation, which can occur in B cells (Reusch et al. 2015), induces a G0/G1 cell cycle arrest (Flemington 2001) could explain the observed AID-dependent hypermutation in the oriP region, as this region will be found as ssDNA only during viral DNA replication. The fact that we did not found evidence for AID-dependent hypermutation in the two oriLyts sequences is consistent with the under-representation of the WRC hotspots.

In our analysis, we found that mutations caused by A3G will result in nonsynonymous amino acid changes in the Gly–Ala repeat of the EBV EBNA-1 latency-associated protein. This region of the protein has a high codon usage bias (GCA for Ala and GGR for Gly), and studies have shown that this repeat region is critical to inhibit antigen presentation through the MHC class 1 pathway (Levitskaya et al. 1997; Apcher et al. 2010). It seems that amino acid changes are not tolerated in this region (Apcher et al. 2010), and therefore A3G activity (with its processive and cooperative action; Chelico et al. 2006; Chaurasiya et al. 2014) in this region could have a strongly detrimental effect for the virus, although given the predominantly cytoplasmic localization of A3G (Lackey et al. 2013), this may occur with low probability.

In conclusion, our results suggest that, even though many EBV and KSHV genes have evolved to limit the action of the AID/A3 enzymes, especially in WRC (AID) and TC motifs, these viral genomes are indeed subjected to AID- and A3-dependent hypermutation. The high under-representation of TC motifs and the presence of TC-dependent mutations is consistent with the fact that this enzyme has a nuclear localization, making it more likely that viral DNA is exposed. However, as stated previously, we cannot ignore that other A3 enzymes other than A3B can be responsible for deamination at TC motifs (Salter, Bennett, and Smith 2016; Salter and Smith, 2018). The existence of specific viral mechanisms to reduce AID activity may have evolved in response to high AID expression in germinal center B cells, and may also explain why we observe only neutral (rather than under-)representation of AID motifs. The number of A3-dependent mutations observed raises the possibility of counteraction mechanisms for these proteins. It will be particularly interesting to determine if the viruses have counter-defense mechanisms against TC-targeting A3s since our analysis showed that many mutations at TC hotspots would be nonsynonymous. Our analysis of hypermutation may well be underestimating the number of mutations since the method we employed to determine the hypermutation relies on a relatively limited number of genomes against a reference sequence. The protein variants that could result from AID/A3 activity may be important, not only for viral fitness, but also for vaccines (epitope determination) and cancer development.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Conflict of interest: None declared.

Acknowledgements

This work was supported by the State University of New York at Stony Brook NIH grant K12-GM-102778 to T.Mar., grants NIH R01AI113134, NIH R41AI115834, the Children's Miracle Network, and funds from the University of Florida to S.B.-M., and NIH R01GM111741 and NIH R01AI132507 to T. Mac. The funders had no role in study design, data collection, and interpretation or the decision to submit the work for publication. We thank Catherine Tang for creating the mutation frequency and distribution R script. We would also like to thank the reviewers for their helpful comments and suggestions.

References

Abascal, F., Zardoya, R., and Telford, M. J. (2010) 'TranslatorX: Multiple Alignment of Nucleotide Sequences Guided by Amino Acid Translations', *Nucleic Acids Research*, 38: 7–13.

Adolph, M. B. et al. (2017) 'Cytidine Deaminase Efficiency of the Lentiviral Viral Restriction Factor APOBEC3C Correlates with Dimerization', *Nucleic Acids Research*, 45: 3378–94.

Apcher, S. et al. (2010) 'Epstein Barr Virus-encoded EBNA1 Interference with MHC Class I Antigen Presentation Reveals a Close Correlation between mRNA Translation Initiation and Antigen Presentation', *PLoS Pathogens*, 6: 1–14.

Bekerman, E. et al. (2013) 'A Role for Host Activation-induced Cytidine Deaminase in Innate Immune Defense against KSHV', *PLoS Pathogens*, 9: e1003748.

Bentz, G. L., Shackelford, J., and Pagano, J. S. (2012) 'Epstein–Barr Virus Latent Membrane Protein 1 Regulates the Function of Interferon Regulatory Factor 7 by Inducing Its Sumoylation', *Journal of Virology*, 86: 12251–61.

Bhagwat, A. S. et al. (2016) 'Strand-biased Cytosine Deamination at the Replication Fork Causes Cytosine to Thymine Mutations in *Escherichia coli*', *Proceedings of the National Academy of Sciences*, 113: 2176–81.

Bishop, K. N. et al. (2004) 'Cytidine Deamination of Retroviral DNA by Diverse APOBEC Proteins', *Current Biology*, 14: 1392–6.

Bogerd, H. P. et al. (2006) 'Cellular Inhibitors of Long Interspersed Element 1 and Alu Retrotransposition', *Proceedings of the National Academy of Sciences of the United States of America*, 103: 8780–5.

Bonvin, M. et al. (2006) 'Interferon-inducible Expression of APOBEC3 Editing Enzymes in Human Hepatocytes and Inhibition of Hepatitis B Virus Replication', *Hepatology*, 43: 1364–74.

Borchert, G. M., Holton, N. W., and Larson, E. D. (2011) 'Repression of Human Activation Induced Cytidine Deaminase by miR-93 and miR-155', *BMC Cancer*, 11: 347.

Chaurasiya, K. R. et al. (2014) 'Oligomerization Transforms Human APOBEC3G from an Efficient Enzyme to a Slowly Dissociating Nucleic Acid-binding Protein', *Nature Chemistry*, 6: 28–33.

Chelico, L. et al. (2006) 'APOBEC3G DNA Deaminase Acts Processively 3′→5′ on Single-stranded DNA', *Nature Structural & Molecular Biology*, 13: 392–9.

——, Pham, P., and Goodman, M. F. (2009) 'Stochastic Properties of Processive Cytidine DNA Deaminases AID and APOBEC3G', *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364: 583–93.

Chen, J., and MacCarthy, T. (2017) 'The Preferred Nucleotide Contexts of the AID/APOBEC Cytidine Deaminases Have Differential Effects When Mutating Retrotransposon and Virus Sequences Compared to Host Genes', *PLOS Computational Biology*, 13: e1005471.

Cheng, A. Z. et al. (2019) 'Epstein–Barr Virus BORF2 Inhibits Cellular APOBEC3B to Preserve Viral Genome Integrity', *Nature Microbiology*, 4: 78–88.

Conticello, S. G. (2008) 'The AID/APOBEC Family of Nucleic Acid Mutators', *Genome Biology*, 9: 229.

Cruz-Muñoz, M. E., and Fuentes-Pananá, E. M. (2017) 'Beta and Gamma Human Herpesviruses: Agonistic and Antagonistic Interactions with the Host Immune System', *Frontiers in Microbiology*, 8: 2521.

Cuevas, J. M. et al. (2015) 'Extremely High Mutation Rate of HIV-1 in Vivo', *PLoS Biology*, 13: e1002251.

Dhar, V., and Schildkraut, C. L. (1991) 'Role of EBNA-1 in Arresting Replication Forks at the Epstein–Barr Virus oriP Family of Tandem Repeats', *Molecular and Cellular Biology*, 11: 6268–78.

Dheekollu, J. et al. (2013) 'Timeless-Dependent DNA Replication-Coupled Recombination Promotes Kaposi's Sarcoma-associated Herpesvirus Episome Maintenance and Terminal Repeat Stability', *Journal of Virology*, 87: 3699–709.

Djavadian, R., Chiu, Y.-F., and Johannsen, E. (2016) 'An Epstein–Barr Virus-encoded Protein Complex Requires an Origin of Lytic Replication in Cis to Mediate Late Gene Transcription', *PLoS Pathogens*, 12: e1005718.

Dorsett, Y. et al. (2008) 'MicroRNA-155 Suppresses Activation-Induced Cytidine Deaminase-Mediated Myc-Igh Translocation', *Immunity*, 28: 630–8.

Dourmishev, L. A. et al. (2003) 'Molecular Genetics of Kaposi's Sarcoma-associated Herpesvirus (human Herpesvirus 8) epidemiology and Pathogenesis', *Microbiology and Molecular Biology Reviews*, 67: 175–212.

Dresang, L. R., Vereide, D. T., and Sugden, B. (2009) 'Identifying Sites Bound by Epstein–Barr Virus Nuclear Antigen 1 (EBNA1) in the Human Genome: Defining a Position-Weighted Matrix to Predict Sites Bound by EBNA1 in Viral Genomes', *Journal of Virology*, 83: 2930–40.

Edwards, R. H. et al. (2004) 'Potential Selection of LMP1 Variants in Nasopharyngeal Carcinoma', *Journal of Virology*, 78: 868–81.

El-Guindy, A., Heston, L., and Miller, G. (2010) 'A Subset of Replication Proteins Enhances Origin Recognition and Lytic Replication by the Epstein–Barr Virus ZEBRA Protein', *PLoS Pathogens*, 6: 53–4.

Epeldegui, M. et al. (2007) 'Infection of Human B Cells with Epstein–Barr Virus Results in the Expression of Somatic Hypermutation-inducing Molecules and in the Accrual of Oncogene Mutations', *Molecular Immunology*, 44: 934–42.

Farina, A. et al. (2005) 'BFRF1 of Epstein–Barr Virus Is Essential for Efficient Primary Viral Development and Egress', *Journal of Virology*, 79: 3703–12.

Fields, B. N., Knipe, D. M., and Howley, P. M. (2013) 'Fields Virology', in D. M., Knipe et al. (eds.) *Fields Virology*, 6th edn. Philadelphia: Lippincott Williams & Wilkins, pp.1898-1959; 2080-2128, https://doi.org/9781451105636

Flemington, E. K. (2001) 'Herpesvirus Lytic Replication and the Cell Cycle: Arresting New Developments', *Journal of Virology*, 75: 4475–81.

Gu, J. et al. (2016) 'Biochemical Characterization of APOBEC3H Variants: Implications for Their HIV-1 Restriction Activity and mC Modification', *Journal of Molecular Biology*, 428: 4626–38.

Hammerschmidt, W., and Sugden, B. (2013) 'Replication of Epstein–Barr Viral DNA', *Cold Spring Harbor Perspectives in Biology*, 5: a0130.

Harris, R. S., and Dudley, J. P. (2015) 'APOBECs and Virus Restriction', *Virology*, 479–480: 131–45.

Heilmann, A. M. F., Calderwood, M. a., and Johannsen, E. (2010) 'Epstein–Barr Virus LF2 Protein Regulates Viral Replication by Altering Rta Subcellular Localization', *Journal of Virology*, 84: 9920–31.

Hippocrate, A., Oussaief, L., and Joab, I. (2011) 'Possible Role of EBV in Breast Cancer and Other Unusually EBV-associated Cancers', *Cancer Letters*, 305: 144–9.

Holden, L. G. et al. (2008) 'Crystal Structure of the Anti-viral APOBEC3G Catalytic Domain and Functional Implications', *Nature*, 456: 121–4.

Hoopes, J. I. et al. (2016) 'APOBEC3A and APOBEC3B Preferentially Deaminate the Lagging Strand Template During DNA Replication', *Cell Reports*, 14: 1273–82.

Hultquist, J. F. et al. (2011) 'Human and Rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H Demonstrate a Conserved Capacity to Restrict Vif-Deficient HIV-1', *Journal of Virology*, 85: 11220–34.

Ito, F. et al. (2017) 'Family-wide Comparative Analysis of Cytidine and Methylcytidine Deamination by Eleven Human APOBEC Proteins', *Journal of Molecular Biology*, 429: 1787–99.

Jemal, A., Bray, F., and Ferlay, J. (1999) 'Global Cancer Statistics: 2011', *CA: A Cancer Journal for Clinicians*, 49: 1.33–64.

Kalchschmidt, J. S. et al. (2016) 'Epstein–Barr Virus Nuclear Protein EBNA3C Directly Induces Expression of AID and Somatic Mutations in B Cells', *The Journal of Experimental Medicine*, 213: 921–8.

Kohli, R. M. et al. (2009) 'A Portable Hot Spot Recognition Loop Transfers Sequence Preferences from APOBEC Family Members to Activation-induced Cytidine Deaminase', *Journal of Biological Chemistry*, 284: 22898–904.

Lackey, L. et al. (2012) 'APOBEC3B and AID Have Similar Nuclear Import Mechanisms', *Journal of Molecular Biology*, 419: 301–14.

—— et al. (2013) 'Subcellular Localization of the APOBEC3 Proteins During Mitosis and Implications for Genomic DNA Deamination', *Cell Cycle*, 12: 762–72.

Larson, E. D., and Maizels, N. (2004) 'Transcription-Coupled Mutagenesis by the DNA Deaminase AID', *Genome Biology*, 5: 211.

Levitskaya, J. et al. (1995) 'Inhibition of Antigen Processing by the Internal Repeat Region of the Epstein–Barr Virus Nuclear Antigen-1', *Nature*, 375: 685–8.

—— et al. (1997) 'Inhibition of Ubiquitin/Proteasome-Dependent Protein Degradation by the Gly–Ala Repeat Domain of the Epstein–Barr Virus Nuclear Antigen 1', *Proceedings of the National Academy of Sciences of the United States of America*, 94: 12616–21.

Li, M. M. H., and Emerman, M. (2011) 'Polymorphism in Human APOBEC3H Affects a Phenotype Dominant for Subcellular Localization and Antiviral Activity', *Journal of Virology*, 85: 8197–207.

Linnstaedt, S. D. et al. (2010) 'Virally Induced Cellular microRNA miR-155 Plays a Key Role in B-cell Immortalization by Epstein–Barr Virus', *Journal of Virology*, 84: 11670–8.

Lu, C. C. et al. (2006) 'Genome-wide Transcription Program and Expression of the Rta Responsive Gene of Epstein–Barr Virus', *Virology*, 345: 358–72.

Madrid, A. S., and Ganem, D. (2012) 'Kaposi's Sarcoma-associated Herpesvirus ORF54/dUTPase Downregulates a Ligand for the NK Activating Receptor NKp44', *Journal of Virology*, 86: 8693–704.

Matsen, F. A., IV et al. (2014) 'A Novel Bayesian Method for Detection of APOBEC3-mediated Hypermutation and Its Application to Zoonotic Transmission of Simian Foamy Viruses', *PLoS Computational Biology*, 10: e1003493.

Minkah, N. et al. (2014) 'Host Restriction of Murine Gammaherpesvirus 68 Replication by Human APOBEC3 Cytidine Deaminases but Not Murine APOBEC3', *Virology*, 454–455: 215–226.

Monajemi, M. et al. (2012) 'Emerging Complexities of APOBEC3G Action on Immunity and Viral Fitness during HIV Infection and Treatment', *Retrovirology*, 9: 35.

—— et al. (2014) 'Positioning of APOBEC3G/F Mutational Hotspots in the Human Immunodeficiency Virus Genome Favors Reduced Recognition by CD8+ T Cells', *PLoS One*, 9: e93428.

Morgan, H. D. et al. (2004) 'Activation-induced Cytidine Deaminase Deaminates 5-Methylcytosine in DNA and Is Expressed in Pluripotent Tissues: Implications for Epigenetic Reprogramming', *Journal of Biological Chemistry*, 279: 52353–60.

Muckenfuss, H. et al. (2006) 'APOBEC3 Proteins Inhibit Human LINE-1 Retrotransposition', *Journal of Biological Chemistry*, 281: 22161–72.

Muramatsu, M. et al. (2000) 'Class Switch Recombination and Hypermutation Require Activation-induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme', *Cell*, 102: 553–63.

Murata, T. et al. (2012) 'Epigenetic Histone Modification of Epstein–Barr Virus BZLF1 Promoter During Latency and Reactivation in Raji Cells', *Journal of Virology*, 86: 4752–61.

Namba-Fukuyo, H. et al. (2016) 'TET2 Functions as a Resistance Factor against DNA Methylation Acquisition During Epstein–Barr Virus Infection', *Oncotarget*, 7: 81512–26.

Norio, P., and Schildkraut, C. L. (2004) 'Plasticity of DNA Replication Initiation in Epstein–Barr Virus Episomes', *PLoS Biology*, 2: 816–33.

Orem, J. et al. (2007) 'Burkitt's Lymphoma in Africa, a Review of the Epidemiology and Etiology', *African Health Sciences*, 7: 166–75.

Pak, V. et al. (2011) 'The Role of Amino-Terminal Sequences in Cellular Localization and Antiviral Activity of APOBEC3B', *Journal of Virology*, 85: 8538–47.

Pavlova, S. et al. (2013) 'An Epstein–Barr Virus Mutant Produces Immunogenic Defective Particles Devoid of Viral DNA', *Journal of Virology*, 87: 2011–22.

Peled, J. U. et al. (2008) 'The Biochemistry of Somatic Hypermutation', *Annual Review of Immunology*, 26: 481–511.

Petersen-Mahrt, S. K., Harris, R. S., and Neuberger, M. S. (2002) 'AID Mutates *E. coli* Suggesting a DNA Deamination Mechanism for Antibody Diversification', *Nature*, 418: 99–104.

Pham, P. et al. (2003a) 'Processive AID-Catalysed Cytosine Deamination on Single-Stranded DNA Simulates Somatic Hypermutation', *Nature*, 424: 103–7.

—— et al. (2013b) 'A Biochemical Analysis Linking APOBEC3A to Disparate HIV-1 Restriction and Skin Cancer', *Journal of Biological Chemistry*, 288: 29294–304.

Purushothaman, P., Uppal, T., and Verma, S. C. (2015) 'Molecular Biology of KSHV Lytic Reactivation', *Viruses*, 7: 116–53.

Rebhandl, S. et al. (2015) 'AID/APOBEC Deaminases and Cancer', *Oncoscience*, 2: 320–33.

Renzette, N. et al. (2014) 'Epstein–Barr Virus Latent Membrane Protein 1 Genetic Variability in Peripheral Blood B Cells and Oropharyngeal Fluids', *Journal of Virology*, 88: 3744–55.

Reusch, J. A. et al. (2015) 'Cellular Differentiation Regulator BLIMP1 Induces Epstein–Barr Virus Lytic Reactivation in Epithelial and B Cells by Activating Transcription From Both the R and Z Promoters', *Journal of Virology*, 89: 1731–43.

Riva, G. et al. (2012) 'How I Treat HHV8/KSHV-Related Diseases in Posttransplant Patients', *Blood*, 120: 4150–9.

Sadler, H. A. et al. (2010) 'APOBEC3G Contributes to HIV-1 Variation Through Sublethal Mutagenesis', *Journal of Virology*, 84: 7396–404.

Salamango, D. J. et al. (2018) 'APOBEC3B Nuclear Localization Requires Two Distinct N-Terminal Domain Surfaces', *Journal of Molecular Biology*, 430: 2695–708.

Salter, J. D., Bennett, R. P., and Smith, H. C. (2016) 'The APOBEC Protein Family: United by Structure, Divergent in Function', *Trends in Biochemical Sciences*, 41: 578–94.

——, and Smith, H. C. (2018) 'Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands', *Trends in Biochemical Sciences*, 43: 606–22.

Saraconi, G. et al. (2014) 'The RNA Editing Enzyme APOBEC1 Induces Somatic Mutations and a Compatible Mutational Signature Is Present in Esophageal Adenocarcinomas', *Genome Biology*, 15: 417.

Schepers, A., Pich, D., and Hammerschmidt, W. (1993) 'A Transcription Factor with Homology to the AP-1 Family Links RNA Transcription and DNA Replication in the Lytic Cycle of Epstein–Barr Virus', *The EMBO Journal*, 12: 3921–9.

Shaban, N. M. et al. (2018) 'The Antiviral and Cancer Genomic DNA Deaminase APOBEC3H Is Regulated by an RNA-Mediated Dimerization Mechanism', *Molecular Cell*, 69: 75–86.e9.

Shapiro, M., Meier, S., and MacCarthy, T. (2018) 'The Cytidine Deaminase Under-Representation Reporter (CDUR) as a Tool to Study Evolution of Sequences Under Deaminase Mutational Pressure', *BMC Bioinformatics*, 19: 163.

Shi, K. et al. (2017) 'Structural Basis for Targeted DNA Cytosine Deamination and Mutagenesis by APOBEC3A and APOBEC3B', *Nature Structural & Molecular Biology*, 24: 131–9.

Simmonds, P., Tuplin, A., and Evans, D. J. (2004) 'Detection of Genome-Scale Ordered RNA Structure (GORS) in Genomes of Positive-stranded RNA Viruses: Implications for Virus Evolution and Host Persistence', *RNA (New York, N.Y.)*, 10: 1337–51.

Starrett, G. J. et al. (2016) 'The DNA Cytosine Deaminase APOBEC3H Haplotype I Likely Contributes to Breast and Lung Cancer Mutagenesis', *Nature Communications*, 7: 12918.

Stenglein, M. D., and Harris, R. S. (2006) 'APOBEC3B and APOBEC3F Inhibit L1 Retrotransposition by a DNA Deamination-Independent Mechanism', *Journal of Biological Chemistry*, 281: 16837–41.

——, Matsuo, H., and Harris, R. S. (2008) 'Two Regions Within the Amino-Terminal Half of APOBEC3G Cooperate to Determine Cytoplasmic Localization', *Journal of Virology*, 82: 9591–9.

Suspene, R. et al. (2011) 'Genetic Editing of Herpes Simplex Virus 1 and Epstein–Barr Herpesvirus Genomes by Human APOBEC3 Cytidine Deaminases in Culture and In Vivo', *Journal of Virology*, 85: 7594–602.

Suspène, R. et al. (2013) 'Efficient Deamination of 5-Methylcytidine and 5-Substituted Cytidine Residues in DNA by Human APOBEC3A Cytidine Deaminase', *PLoS One*, 8: e63461.

Tobollik, S. et al. (2006) 'Epstein–Barr Virus Nuclear Antigen 2 Inhibits AID Expression During EBV-Driven B-cell Growth', *Blood*, 108: 3859–64.

Vieira, V. C. et al. (2013) 'The Role of Cytidine Deaminases on Innate Immune Responses against Human Viral Infections', *BioMed Research International*, 2013: 1.

Wang, Q. et al. (2017) 'The Cell Cycle Restricts Activation-Induced Cytidine Deaminase Activity to Early G1', *The Journal of Experimental Medicine*, 214: 49–58.

Watanabe, T. et al. (2015) 'The Epstein–Barr Virus BRRF2 Gene Product Is Involved in Viral Progeny Production', *Virology*, 484: 33–40.

Willems, L., and Gillet, N. A. (2015) 'APOBEC3 Interference During Replication of Viral Genomes', *Viruses*, 7: 2999–3018.

Wood, N. et al. (2009) 'HIV Evolution in Early Infection: Selection Pressures, Patterns of Insertion and Deletion, and the Impact of APOBEC', *PLoS Pathogens*, 5: e1000414.

Wu, L. et al. (2009) 'Epstein–Barr Virus LF2: An Antagonist to Type I Interferon', *Journal of Virology*, 83: 1140–6.

Young, L. S., Fah Yap, L., and Murray, P. G. (2016) 'Epstein–Barr Virus: More than 50 Years Old and Still Providing Surprises', *Nature Reviews Cancer*, 16: 789–802.

Yu, Q. et al. (2004) 'Single-Strand Specificity of APOBEC3G Accounts for Minus-Strand Deamination of the HIV Genome', *Nature Structural & Molecular Biology*, 11: 435–42.

Zhu, X. et al. (2014) 'MiR-155 up-regulation by LMP1 DNA Contributes to Increased Nasopharyngeal Carcinoma Cell Proliferation and Migration', *European Archives of Oto-Rhino-Laryngology*, 271: 1939–45.