PNAS RESEARCH ARTICLE | EVOLUTION OPEN ACCESS

Host ZAP activity correlates with the levels of CpG suppression in primate lentiviruses

Rayhane Nchioua[a,1], Dorota Kmiec[a], Veronika Krchlikova[b], Sarah Mattes[a], Sabrina Noettger[a], Frederic Bibollet-Ruche[c], Ronnie M. Russell[c], Konstantin M. J. Sparrer[a,d], Thomas Charpentier[e], Frédéric Tardy[f], Steven E. Bosinger[g], Daniel Sauter[b], Beatrice H. Hahn[c], and Frank Kirchhoff[a,1]

Affiliations are included on p. 10.

Edited by Malcolm Martin, National Institute of Allergy and Infectious Diseases, Bethesda, MD; received September 23, 2024; accepted February 24, 2025

Zinc-finger antiviral protein (ZAP) is thought to drive the suppression of CpG dinucleotides in many viruses to mimic the composition of their host genomes. However, in vivo evidence is sparse. Here, we investigated the reasons for unusually high CpG levels in SIVmus and SIVmon from mustached and mona monkeys, descendants of one of the precursors of HIV-1. We show that SIVmus is not resistant to ZAP inhibition. Instead, these *Cercopithecus* monkey hosts differ from other primate species by a splice site mutation and express the poorly active extralarge XL rather than the highly active L isoform of ZAP. Similarly, higher CpG levels in endogenous prosimian lentiviruses were associated with low activity of the corresponding host lemur ZAPs. In addition, lemur genes also show lower CpG suppression than other primates. Thus, the antiviral activity of ZAP not only affects suppression of CpG dinucleotides in viral transcripts but possibly also host genomes.

zinc-finger antiviral protein | CpG suppression | primate lentiviruses | virus-host arms race

Significance

Zinc-finger antiviral protein (ZAP) plays a key role in suppressing CpG dinucleotides in viral genomes, allowing viruses to mimic their host's genome composition to evade immune detection. Our research explains why certain monkey viruses, including ancestors of HIV-1, have unusually high CpG levels. In silico analyses suggest that a G-to-A mutation upstream of splice donor 4 promotes expression of a less active ZAP isoform. We also identified five mutations in mustached monkey ZAP that reduce its antiviral activity. Our findings show that ZAP activity influences CpG suppression in primate lentiviruses and possibly their hosts, providing important insights into viral evolution and genome regulation in primates.

Vertebrate species are in a constant arms race with viral pathogens (1). To detect and control viral infections the immune system of higher organisms must be able to distinguish self from non-self (2, 3). This represents a challenge to viruses since they rely on the host's transcription, translation, and metabolic machinery to propagate, with all viral components being derived from the host cell. Pattern recognition receptors (PRRs) achieve self/non-self-discrimination by recognizing RNA or DNA molecules with unusual structure, composition, or localization. To avoid immune recognition and consequently the induction of antiviral defense mechanisms, many viruses evolved to mimic the features of their respective hosts (4–6). Strong reciprocal selective pressures have driven the evolution and hence the features and functions of both host and viral genomes over millions of years (7, 8).

One hallmark of mammalian genomes (9) that is mimicked by many viral pathogens (6, 10) is a paucity of CpG dinucleotides. It has been suggested that a major driver of CpG suppression (i.e. the observed levels of CpG dinucleotides divided by the expected levels of CpG dinucleotides based on the number of C and G residues in the sequence) in viruses is the zinc finger antiviral protein (ZAP), which specifically targets and mediates degradation of CpG-containing RNAs (11). ZAP was initially identified as a broad-acting antiviral effector that destabilizes viral RNAs (12, 13) and restricts numerous viral pathogens, including alphaviruses, filoviruses, hepatitis B virus, influenza A virus, and retroviruses (14–17). In addition, ZAP may also act as a PRR for CpG-containing nucleic acids as it potentiates the host's antiviral response (18, 19). ZAP has minimal effects on human mRNAs, presumably because CpG dinucleotides are sparse in human transcripts (20, 21). It is believed that the loss of CpG dinucleotides in vertebrate genomes is the result of C-to-T mutations introduced by CpG-specific DNA methyl transferases and methyl-cytosine deamination over millions of years (22, 23). Host CpG suppression is mirrored by many RNA viruses infecting them (6, 10).

It has been proposed that long-term evolutionary pressures exerted by the antiviral activity of ZAP have driven CpG suppression in viral genomes (24). However, there is no direct evidence for this hypothesis. Primate lentiviruses have been detected in more than 40 Old World nonhuman primate species and are well studied as some have given rise to HIV (25, 26). HIV-1, the main cause of AIDS has a complex evolutionary history involving cross-species transmission and recombination events of primate lentiviruses from monkeys to great apes to humans (27). Interestingly, the levels of CpG suppression in primate lentiviruses that ultimately gave rise to HIV-1 vary substantially (28). SIVcpz from chimpanzees is the direct precursor of pandemic HIV-1 strains (29) and evolved from a precursor of viruses infecting greater spot-nosed (SIVgsn), mona (SIVmon), and mustached (SIVmus)

Author contributions: R.N., D.K., V.K., F.B.-R., K.M.J.S., S.E.B., B.H.H., and F.K. designed research; R.N., D.K., V.K., S.M., S.N., F.B.-R., R.M.R., and S.E.B. performed research; V.K., F.B.-R., T.C., F.T., S.E.B., D.S., and B.H.H. contributed new reagents/analytic tools; R.N., D.K., V.K., F.B.-R., K.M.J.S., S.E.B., D.S., B.H.H., and F.K. analyzed data; and B.H.H. and F.K. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: nchioua.rayhane@uni-ulm.de or frank.kirchhoff@uni-ulm.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2419489122/-/DCSupplemental.

Published April 3, 2025.

monkeys that had recombined with a precursor of current SIVrcm from red-capped mangabeys or SIVmnd2 found in mandrills (30, 31). SIVcpz was transmitted to gorillas and humans, giving rise to SIVgor and HIV-1 groups M and N, respectively (25, 27), while HIV-1 groups O and P are the result of two zoonotic transmission events of SIVgor (32). The genomes of currently circulating SIVgsn, SIVmon, and SIVmus infecting greater spot-nosed (GSN), mona, and mustached monkeys, respectively, contain significantly higher levels of CpGs than those of other primate lentiviruses including HIV-1, SIVcpz, and SIVgor (28).

Here, we generated an infectious molecular clone of SIVmus and show that its reduced CpG suppression is not due to viral resistance to ZAP inhibition. Instead, ZAPs from mustached and mona monkeys display poor antiviral activity compared to ZAP orthologs from other primate species, including great apes and humans. Following up on the relatively high CpG content of endogenous prosimian lentiviruses, we found that lemurs also express poorly active ZAPs. Unlike in GSN, mona, and mustached monkeys, both the genomes of the lemur hosts and the endogenous lentiviruses that infiltrated their germ line several Mya (33), are relatively CpG-rich. Thus, the antiviral activity of ZAP in primates appears to not only affect CpG suppression in their lentiviruses, but may also impact host CpG content over long evolutionary periods.

Results

Relatively High CpG Frequency in the SIVgsn/Mus/Mon Lineage. HIV-1 originated from SIVs infecting chimpanzees and gorillas (25, 27). Chimpanzees were most likely coinfected by several SIVs found in monkey species including members of the *Cercopithecidae* family that recombined to generate SIVcpz (Fig. 1*A*) (30, 31).

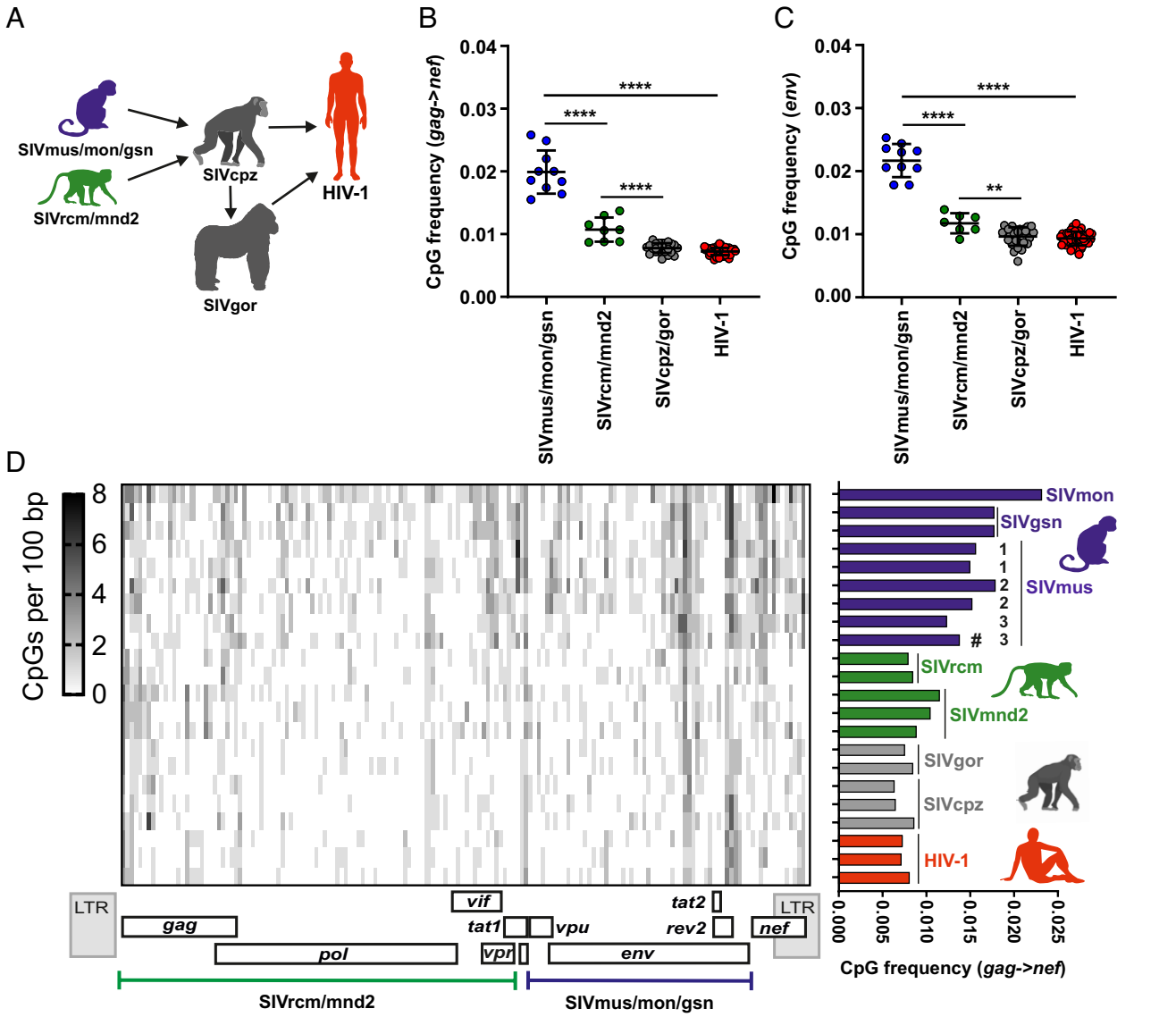

Fig. 1. Relatively high CpG levels in SIVs infecting MUS, MON, and GSN monkeys. (*A*) Schematic presentation of viral transmission events preceding the emergence of HIV-1. (*B* and *C*) Frequency of CpG dinucleotides in near full-length viral genomes gag to nef coding regions (*B*) or the env gene alone (*C*). Lentiviral sequences were grouped according to their phylogenetic relationships. Statistical analysis was performed using the Mann–Whitney test. Shown is the mean ±SD. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. (*D*) Heatmap showing the frequency of CpG dinucleotides (white = 0 CpGs, black = 9 CpGs) per 100 bp in the indicated aligned lentiviral genomes (see *SI Appendix* for Genbank accession numbers). Relative position of CpGs is shown in reference to the HIV-1 genome. The green and blue lines indicate which parts of the SIVcpz genome originated from precursors of SIV nowadays infecting red-capped mangabeys and mandrills (SIVrcm/mnd2) or closely related Cercopithecus monkeys (SIVmus/mon/gsn), respectively. The *Right* panel shows the frequency of CpG dinucleotides in the lentiviral genomes, i.e. number of CpGs divided by the length of the genomic region. The (#) symbol indicates the SIVmus clone generated in the present study. Three SIVmus lineages (denoted 1, 2, and 3) were used for the CpG analysis.

Despite their common evolutionary history, the frequency of CpG residues in these primate lentiviruses varies. HIV-1, SIVcpz, and SIVgor show frequencies of CpG dinucleotides ranging from 0.6 to 0.8%, while members of the SIVgsn/mus/mon lineage exhibit ~2- to 4-fold higher CpG frequencies (Fig. 1*B*). These differences are highly significant and also observed for the *env* gene (Fig. 1*C*) that determines ZAP sensitivity of HIV-1 (28). Importantly, the higher levels of CpGs in SIVgsn/mus/mon are not limited to specific regions, but observed throughout their genomes (Fig. 1*D*). Since the *env* gene of SIVcpz was derived from a precursor of SIVmus/mon/gsn (Fig. 1*D*), these data indicate a significant drop in CpG frequency after transmission of these SIVs from *Cercopithecus* monkeys to apes and humans.

SIVmus Is Highly Sensitive to ZAP Restriction. Our finding that the viral CpG content decreased after transmission of SIV from *Cercopithecus* monkeys to chimpanzees suggested that either SIVgsn/mus/mon evades or counteracts ZAP restriction or that ZAP has reduced antiviral activity in these monkeys. To address this, we generated a full-length molecular clone of SIVmus infecting *Cercopithecus cephus* (Fig. 2*A*) by synthesizing a published proviral sequence (34, 35) as a representative of the CpG-high SIVgsn/mus/mon group. Sequence analyses confirmed that the SIVmus-11Gab-Pts02 (referred to as SIVmus hereafter) genome was intact and contained a *vpu* gene similar to HIV-1 (Fig. 2*B*). Transfection of HEK293T cells with the proviral SIVmus DNA yielded viral particles capable of infecting TZM-bl reporter cells (Fig. 2*C*), which express high levels of human CD4, CCR5, and CXCR4 (36). To assess whether this molecular clone of SIVmus is replication competent, we infected human peripheral blood mononuclear cells (PBMCs) and CD4+ T cells. SIVmus replicated in both of them, but required VSV-G pseudotyping and treatment with a fibrillar attachment factor (37) for initial infection (Fig. 2*D* and *SI Appendix*, Fig. S1*A*). SIVcpz MB897, which represents a close genetic relative of HIV-1 group M (38), replicated less efficiently in primary human cells than HIV-1 but did not require special treatment for initiation of infection.

To test the susceptibility of SIVmus to ZAP restriction, we determined infectious virus yields after cotransfection of ZAP knockout (KO) HEK293T cells (11) with the SIVmus proviral construct and different doses of a ZAP expression or an empty control vector. As a control, we also examined the HIV-1 NL4-3 and NHG molecular clones and the mutant NHG-L clone that contains 145 synonymous mutations generating 37 additional CpG sites (11). We found that SIVmus was about as sensitive to ZAP restriction as the HIV-1 NHG-L construct containing the artificially high CpG numbers (Fig. 2*E* and *SI Appendix*, Fig. S1*B*). These results suggest that the high CpG content of SIVmus renders it ZAP sensitive, which may contribute to its reduced replication fitness in primary human cells.

Mustached and Mona Monkey Blood Cells Express the XL Rather Than L Isoform of ZAP. Four different isoforms of ZAP that arise from alternative splicing have been reported (39). The two main isoforms in humans are ZAP-L (long) and ZAP-S (short), differing by the absence of the C-terminal PARP-like

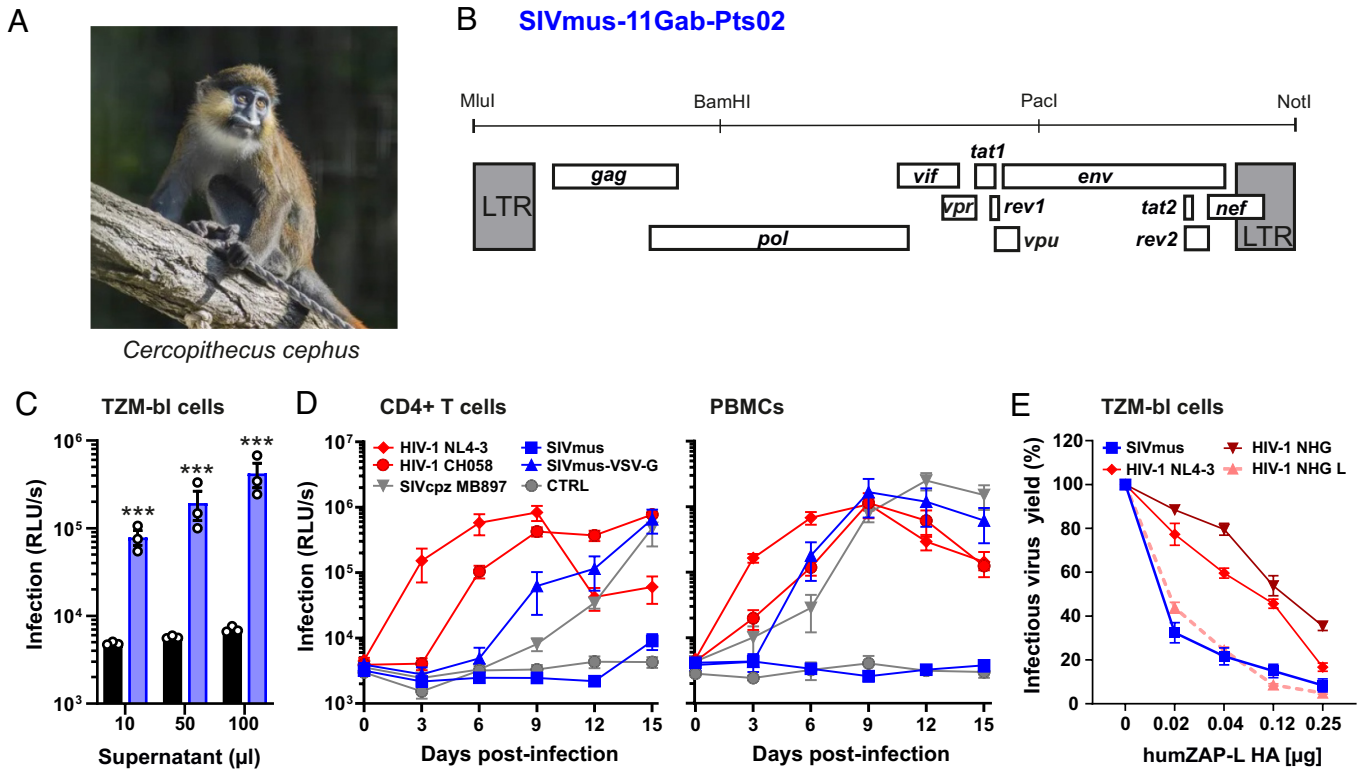

Fig. 2. Susceptibility of SIVmus to ZAP restriction. (*A*) Image of a mustached guenon (Cercopithecus cephus); pictures taken from Zoo Beauval in France, courtesy of Dr. Baptiste Mulot. (*B*) Genomic organization of the molecular clone of SIVmus-11Gab-Pts02(34). (*C*) HEK293T cells were transfected with SIVmus IMC or empty vector control. Supernatants obtained 2 d later were used to infect TZM-bl cells and β-galactosidase activities determined 3 d later. Data present the mean of three independent experiments each tested in triplicates (±SEM). RLU/s, relative light unit per second. Significant differences compared to control using the t test are indicated as ***$P < 0.001$. (*D*) Stimulated PBMCs and CD4+T cells from four different healthy human donors were infected with the indicated viruses including VSV-G pseudotyped SIVmus. Virus stocks were treated with EF-C, a fibrillar attachment factor, to enhance initial infection (37). Supernatants were harvested at 3 d intervals and examined for infectious virus production by TZM-bl reporter assay. Shown are mean values from the four donors (±SEM) each measured in duplicate. (*E*) HEK293T ZAP KO cells were cotransfected with increasing amounts of humZAP-L expression vector and indicated proviral constructs of HIV-1 NL4-3, HIV-1 NHG, and HIV-1 NHG L CpG high mutant. Infectious virus yields determined by TZM-bl assay are provided relative to those obtained in the absence of ZAP set as (100%). Shown are mean values from three to four independent experiments, each tested in technical triplicates (±SEM).

domain from the shorter form (Fig. 3*A*). In humans, ZAP-L exhibits higher antiviral activity than ZAP-S (40, 41), while ZAP-S is more strongly inducible by interferons (19). The two remaining isoforms, ZAP-M (medium) and ZAP-XL (extralong) contain an extended exon 4 and are only expressed at low levels in humans (39). We initially focused on ZAP-L since it exerts the strongest antiviral activity due to its PARP-like domain (40–43), which is lacking in the S and M isoforms (39). To address potential species-specific differences in the antiviral activity of ZAP, we isolated total mRNA from blood samples obtained from different primate species including mustached (MUS) and mona (MON) monkeys. It has been reported that ZAP-L is expressed at substantially higher levels than ZAP-XL in a variety of human cell lines (39). Thus, it came as a surprise that only transcripts encoding ZAP-XL were PCR amplified from MUS and MON samples (Fig. 3*B*). In contrast, amplification products from blood samples of humans (HUM), mandrills (MND), and macaques (MAC) represented ZAP-L transcripts, while samples from chimpanzees (CPZ) and sabaeus monkeys (SAB) yielded both isoforms (Fig. 3*B* and *SI Appendix*, Fig. S2*A*). This was the case despite the fact that the same PCR primer sets were used for all amplifications. The human *ZAP/ZC3HAV1* gene comprises a total of 13 exons, and expression of the L and XL isoforms is determined by alternative splicing at splice donor (SD) sites 4 and 5, respectively (Fig. 3*C*). To predict the relative usage of these splice donors in different primate species, we took advantage of SpliceRover, a deep learning approach using convolutional neural networks reported to outperform the accuracy of earlier splice prediction tools (44). The SpliceRover in silico predictions suggest that a reduced strength of SD4 in MUS monkeys and related *Cercopithecus* species contributes to the expression of the XL rather than L isoform (Fig. 3*D*). In silico mapping revealed that poor SD4 usage is largely the result of a single G to A change upstream of SD4 that is unique to *Cercopithecus* species and (to a lesser extent) a downstream G to A change compared to other monkey species (*SI Appendix*, Fig. S2*B*). These results indicate that

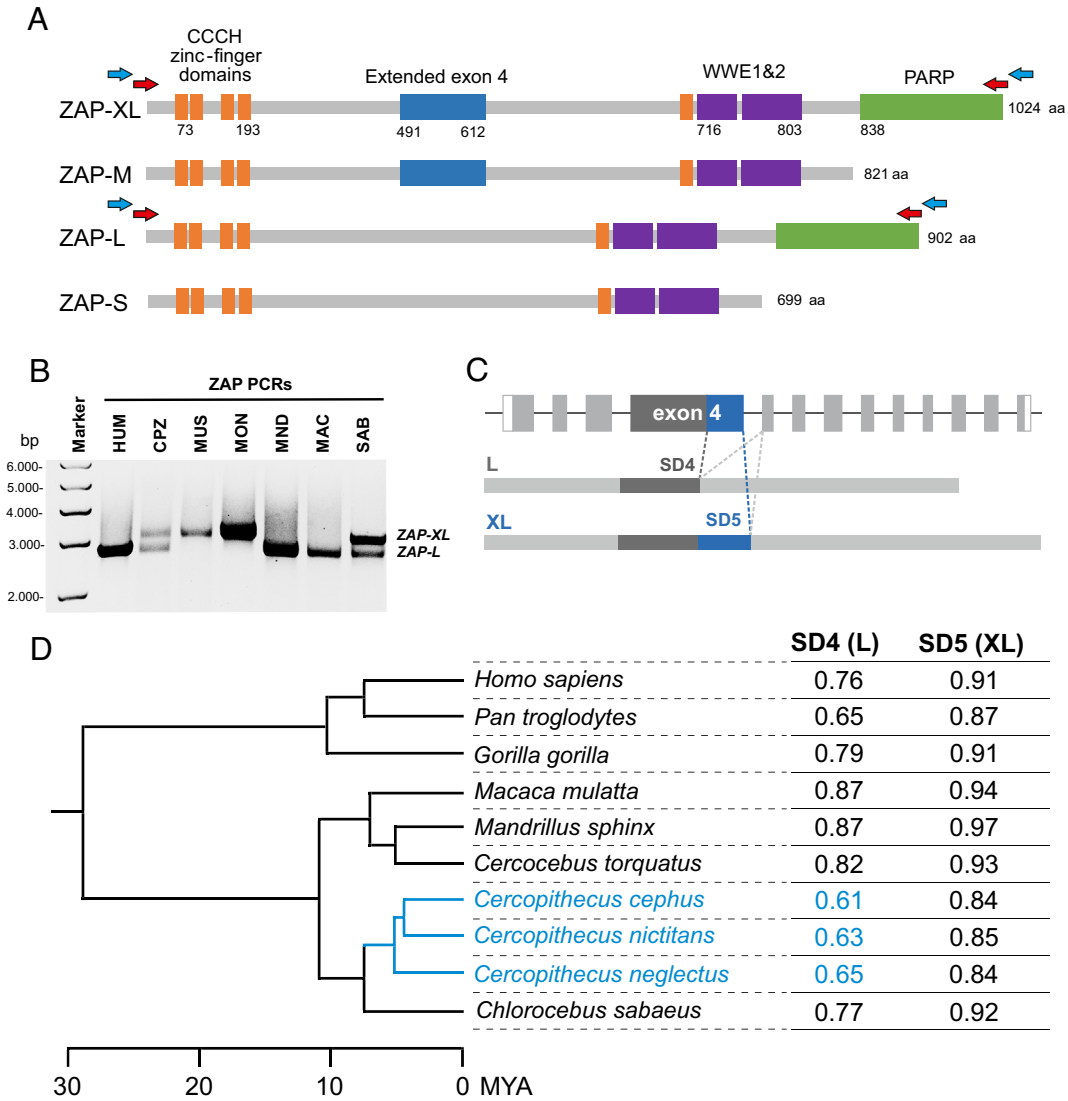

Fig. 3. Species-dependent differences in the expression of ZAP splice variants. (*A*) Domain organization of different ZAP isoforms: four N-terminal CCCH zinc finger RNA binding domains (orange boxes) and two central WWE domains (purple boxes) are shared by all four isoforms. The catalytically inactive C-terminal PARP domain (green box) is shared by ZAP-L and -XL. The ZAP-XL and M isoforms harbor an extended exon 4 (blue box). Domains are not drawn to scale. Primers used for ZAP amplification are indicated by blue and red arrows. (*B*) Amplification of ZAP genes from blood cells derived from the indicated primate species. (*C*) Schematic of the ZAP gene encompassing a total of 13 exons and generation of ZAP-L and ZAP-XL via usage of splice donor 4 (SD4) and 5 (SD5), respectively. (*D*) Splice donor usage for ZAP orthologs of the indicated primate species was predicted using SpliceRover (44). The respective scores are shown on the *Right*. Scores range from 0.0 to 1.0, with high scores indicating high probability of splice donor usage. The phylogenetic tree on the *Left* is based on timetree.org and Shao et al. 2023 (45). *Cercopithecus spec.* are highlighted in blue.

a mutation upstream of SD4 that emerged in a common ancestor of MUS, MON, GSN monkeys, and other *Cercopithecus* species, such as *C. neglectus*, promotes expression of the XL rather than L isoform of ZAP. Altogether, these findings suggest that ZAP-XL is the most abundant antiviral isoform expressed by MUS and MON monkeys, while ZAP-L is expressed in other primate species.

Mustached and Mona Monkey ZAP Show Reduced Antiviral Activity. For functional analyses, we generated constructs expressing HA-tagged versions of the endogenously expressed XL and L isoforms of MUS, MON, and HUM ZAP, respectively. While only ZAP-XL was amplified from MUS and MON PBMCs (Fig. 3*B*), we cannot exclude the possibility that ZAP-L might also be expressed in these monkeys. Thus, we also generated constructs expressing HA-tagged L isoforms of MUS and MON ZAP. Western blot analysis showed that all ZAP variants were expressed in transfected HEK293T cells, albeit the XL isoforms at lower levels than the L variants (Fig. 4*A*). The ZAP-XL orthologs amplified from MUS and MON PBMCs both inhibited wild-type HIV-1 NL4-3, the two transmitted founder CH058 and CH077 strains (46) and SIVmus with substantially lower efficiency than humZAP-L (Fig. 4*B* and *SI Appendix*, Fig. S3). SIVmus was significantly more sensitive than HIV-1 to restriction by humZAP-L and monZAP-L at the lowest concentration of expression vectors and to the weaker ZAP-XL orthologs at all concentrations used (*SI Appendix*, Fig. S3). The artificially generated musZAP-L isoform was also less active than humZAP-L, while artificial monZAP-L was almost as active as humZAP-L (Fig. 4*B* and *SI Appendix*, Fig. S3). In contrast, authentic amplified monZAP-XL showed the lowest antiviral activity of all ZAP variants analyzed. Dose-dependent analyses

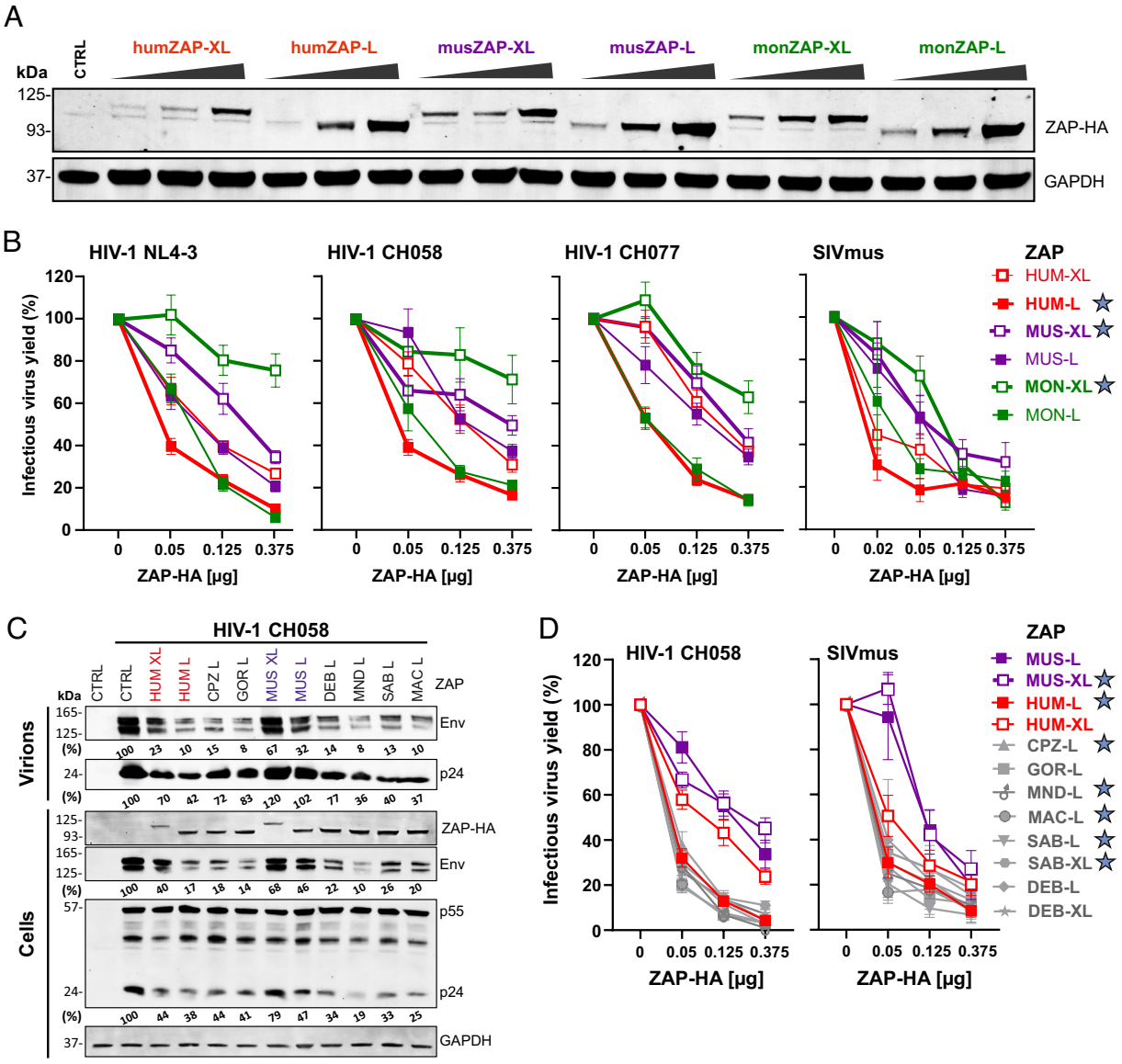

Fig. 4. Antiviral activity of primate ZAPs. (*A*) HEK293T ZAP KO cells were cotransfected with increasing amounts of vectors expressing the indicated N-terminally HA-tagged ZAP isoforms or empty vector control (CTRL) and analyzed by western blot. (*B*). Inhibition of HIV-1 and SIVmus by increasing amounts of ZAP orthologs, amplified from HUM, MUS, or MON blood or chemically synthesized ZAP genes. Stars indicate ZAP isoforms amplified from primate PBMCs, while their absence indicates synthesized isoforms. Shown are means of three independent experiments, each tested in duplicates (SIVmus) or triplicates (HIV-1), ± SEM. (*C*) Impact of ZAP on HIV-1 protein expression. HEK293T ZAP KO cells were cotransfected with 0.05 µg of indicated ZAP constructs and HIV-1 CH058 provirus. Whole-cell lysates and supernatants were analyzed for ZAP and viral proteins expression levels by western blot. Values shown under the blot represent normalized Env and p24 levels, adjusted to the internal control GAPDH and normalized to the no-ZAP condition (100%) (*D*) HEK293T ZAP KO cells were cotransfected with CH058 (*Left* panel) or SIVmus (*Right* panel) proviral constructs and increasing amounts of vectors expressing the ZAP-XL or ZAP-L isoforms. Infectious virus yield values were normalized to no ZAP set as (100%). ZAP orthologs amplified from primate blood are indicated with a star symbol. Shown are mean values (± SEM) obtained from three to eight independent experiments, each tested in triplicate.

allowed us to compare the effects of ZAP-XL and ZAP-L at similar expression levels, i. e. 0.375 μg and 0.125 μg expression vectors, respectively (Fig. 4*A*). We found that humZAP-L is moderately more active than humZAP-XL, while both musZAP-L and musZAP-XL show poor activity and monZAP-XL is much less active than monZAP-L (Fig. 4*B*). Thus, dependent on the species origin both reduced expression and function contribute to the reduced antiviral activity of the ZAP-XL isoform.

To further assess potential attenuation of musZAP function, we analyzed the antiviral activity of ZAP-L isoforms from six additional primate species. Chimpanzee (CPZ, *Pan troglodytes*), Rhesus macaque (*Macaca mulatta*), green monkey, also known as sabaeus monkey (SAB, *Chlorocebus sabaeus*) and mandrill (MND, *Mandrillus sphinx*) *ZAP* orthologs were amplified from PBMCs. In addition, we synthesized the ZAP-L coding sequences of gorillas (GOR, *Gorilla gorilla*), as well as the L and XL isoforms of De Brazza monkeys (DEB, *Cercopithecus neglectus*) based on published sequences (40). The great apes were selected because they are the original source of HIV-1. DEB was chosen since it is genetically closely related to MUS and MON monkeys but harbors an SIV strain (SIVdeb) that is strongly CpG suppressed. All monkey, great ape, and human ZAP-L isoforms (18, 39, 41, 47) were expressed and reduced HIV-1 Env and p24 capsid expression in transfected HEK293T ZAP KO cells (Fig. 4*C*). MusZAP-L and -XL proteins inhibited HIV-1 CH058 and SIVmus with lower efficiency compared to ZAPs from all other primate species analyzed (Fig. 4*D*). All *ZAP* orthologs that showed impaired activity in reducing infectious virus yields also had lesser effects on viral protein levels in cell culture extracts and supernatants (*SI Appendix*, Fig. S4). *ZAP-L* and *-XL* isoforms were both obtained from SAB PBMCs (Fig. 3*B*), while we were unable to PCR amplify *ZAP* from DEB PBMCs. Importantly, ZAP-XL proteins of SAB and DEB show four consistent amino acid changes compared to musZAP-XL (*SI Appendix*, Fig. S5) and were substantially more active against HIV-1 and SIVmus (Fig. 4*D*). Altogether, our results indicated that preferential expression of the XL isoform and specific amino acid changes attenuate the antiviral activity of MUS and MON ZAP.

While previous studies generally reported that humZAP-L is more active than humZAP-S, the latter can still display significant antiviral activity (18, 39, 41, 47). Our PCR methods specifically amplified the larger L and XL isoforms of ZAP from PBMC samples. Specific amplification of the shorter forms M and S is challenging because their entire sequence is included within the longer isoforms. To exclude the possibility that the S or M forms of musZAP are highly antivirally active, we generated them by site-directed mutagenesis from the L and XL encoding constructs, respectively. We also generated these forms for human, mustached, mona, and De Brazza monkey ZAP for comparison. All four isoforms of musZAP showed poor antiviral activity (*SI Appendix*, Fig. S6*A*). As expected, the L isoforms of humZAP and debZAP displayed higher antiviral activity compared to the S isoforms, although the effect saturated at the highest dose of ZAP expression vectors (*SI Appendix*, Fig. S6*A*). The synthesized S and L forms of monZAP showed similar activity at higher doses. However, both of them were substantially less active compared to HUM and DEB ZAP-L (*SI Appendix*, Fig. S6*B*). As outlined above, only the XL but not the L isoform was amplified from MUS and MON PBMCs (Fig. 3*B*). Importantly, all M and XL isoforms of musZAP and monZAP, which contain the extended exon 4, exhibited very poor antiviral activity (*SI Appendix*, Fig. S6*A*), although they were efficiently expressed as shown by Western blot analysis (*SI Appendix*, Fig. S6*B*). ZAP alleles that strongly inhibited infectious virus production also efficiently impaired viral protein expression (*SI Appendix*, Fig. S6*C*). Altogether, the data show that all musZAP isoforms, as well as endogenously expressed monZAP-XL and the shorter M isoform, exhibit poor antiviral activity.

A Combination of Five Mutations Restores Antiviral Activity of musZAP. The deduced amino acid sequence of musZAP-XL differed in a total of 111 positions from humZAP-XL and only by four changes from monZAP-XL. Sequence alignments revealed that musZAP differs at five amino acid positions from fully antivirally active ZAP orthologs (*SI Appendix*, Fig. S7). To identify determinants of musZAP attenuation, we introduced mutations of IH120/123VN, T306A, M457I, and A604T (A726T in ZAP-XL) individually and in combination into the musZAP-L and -XL orthologs and the reverse changes into humZAP-L (Fig. 5*A*). Notably, these ZAP residues differ from those shown to be under positive selection pressure (*SI Appendix*, Fig. S7) (47). Western blot analyses showed that all mutant ZAP variants were efficiently expressed (Fig. 5*B* and *SI Appendix*, Fig. S8*A*). Moreover, individual and combined mutations of VN120/123IH, A306T, I457M, and T604A had no significant effects on the antiviral activity of humZAP-L (Fig. 5*C*). However, the combined changes significantly increased the activity of both musZAP-L and -XL in restricting HIV-1 and SIVmus (Fig. 5 *D* and *E* and *SI Appendix*, Fig. S8*A*). Individual mutations in musZAP resulted in intermediate phenotypes, with T306A showing the most pronounced effect. Sequence analyses revealed that the I457M mutation in ZAP-L is specific for mustached and mona monkeys, while V120I is also found in lemurs and lorises. N123H is found in other *Cercopithecus* monkeys and mandrills, and A306T is found in New World monkeys but not in any great ape or Old World primate species analyzed (*SI Appendix*, Table S1). Thus, the sequence variations found in ZAPs of mustached and mona monkeys are unusual for primate lentiviral hosts. However, since they did not impair the antiviral activity of humZAP, they do not allow to define loss of functional activity.

To determine whether the activity-enhancing mutations in musZAP affect interaction with the ZAP cofactors TRIM25 (48, 49) and KHNYN (50), we performed coimmunoprecipitation (Co-IP) assays. Neither HUM nor MUS ZAPs and their respective mutants differed significantly in their ability to precipitate endogenous HUM KHNYN and TRIM25 (*SI Appendix*, Fig. S8*B* and *C*). These results show that both expression of the XL isoform and a combination of changes throughout the protein are responsible for the low antiviral activity of musZAP.

Impaired ZAP Function and Relatively High CpG Levels in Lemurs. Our results indicated that the levels of lentiviral CpG suppression might be an indicator for antiviral ZAP activity in the respective host species. To further address this, we examined CpG suppression levels in lentiviruses infecting various species, including cats, sheep, goats, horses, cattle, and lemurs (Fig. 6*A*). These analyses revealed that pSIVgml, an endogenous lentivirus that has been detected in the genome of the gray mouse lemur (GML, *Microcebus murinus*) (51), shows low levels of CpG suppression. Similarly, among known simian β-retroviruses, prosimian retrovirus 1 infecting lemur Coquerel's sifaka (CSL, *Propithecus coquereli*) showed the highest CpG frequency (*SI Appendix*, Fig. S9). To investigate whether higher levels of viral CpGs are an indication of less active ZAP in the respective host species, we synthesized the genes encoding the ZAP-L proteins of GML and CSL. The in silico translated amino acid sequences of gmlZAP and cslZAP show only ~60% amino acid identity to humZAP and contain some deletions (*SI Appendix*, Fig. S10). Both lemur ZAP-L proteins were detectable in transfected HEK293T ZAP KO cells (Fig. 6*B*). However, in agreement with

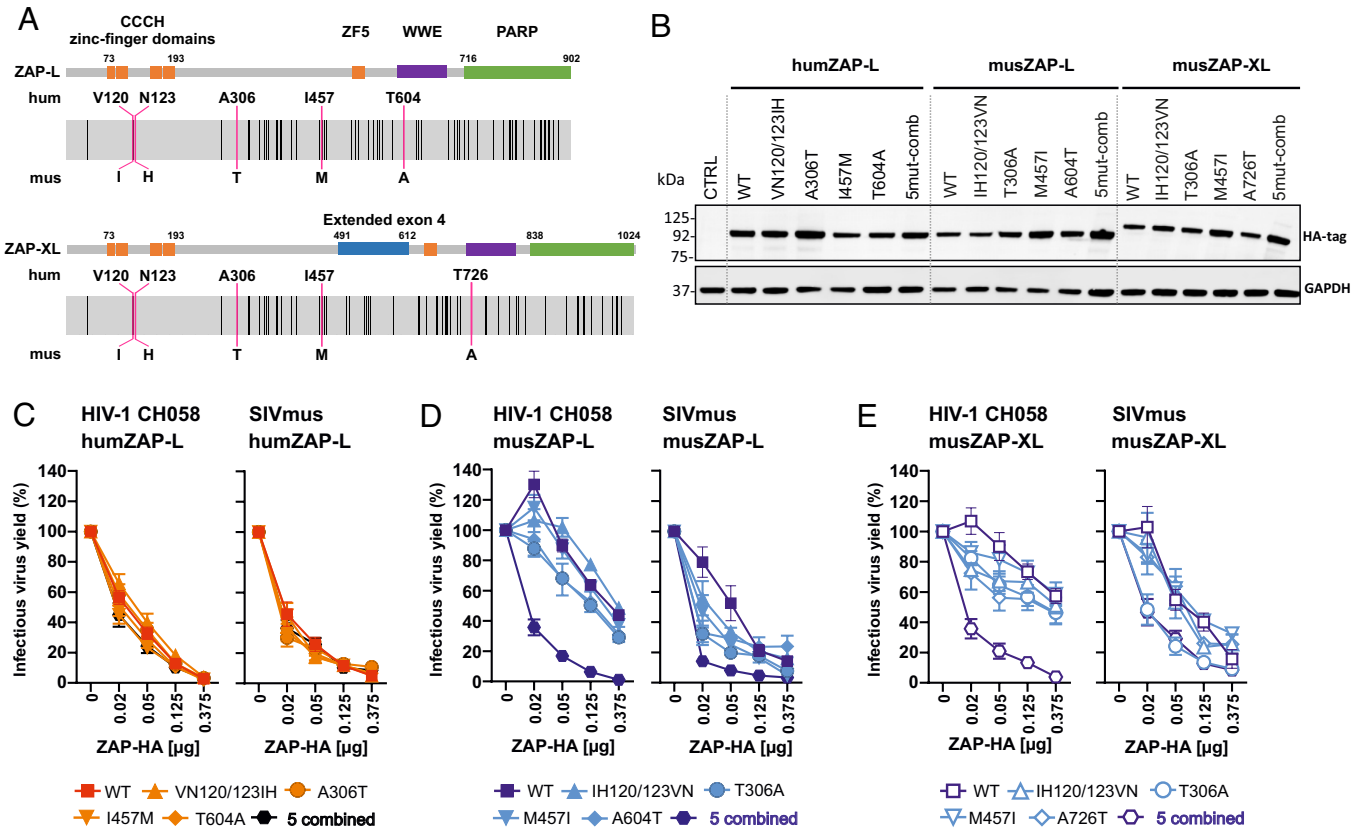

Fig. 5. A combination of five mutations restores musZAP antiviral activity. (*A*) Schematic presentation of mutations in HUM and MUS ZAPs analyzed. Amino acid residues selected for functional analyses are highlighted in pink. Black lines indicate amino acid differences between HUM and MUS ZAPs. (*B*) Representative western blot of whole cell lysates showing expression levels of parental and mutant ZAPs. (*C*–*E*) HEK293T ZAP KO cells were cotransfected with proviral HIV-1 CH058 or SIVmus constructs and increasing amounts of vectors expressing the N-terminally HA-tagged isoforms of wt or mutant humZAP-L (*C*), musZAP-L (*D*), or mus ZAP-XL (*E*). Infectious virus yield was determined using TZM-bl assay. For each proviral construct, values were normalized to the infectious virus yield obtained in the absence of ZAP (100%). Shown are the mean (± SEM) of three to four independent experiments each performed in triplicates.

its smaller size (*SI Appendix*, Fig. S10), cslZAP-L migrated faster in SDS-PAGE and was only detectable at low levels. Similar to musZAP-L, the GML and CSL ZAP-L proteins showed lower antiviral activity than humZAP-L (Fig. 6*C*). pSIVgml entered the germline of mouse lemurs several Mya (51) and the proviral DNA has since then evolved together with the host genome. We

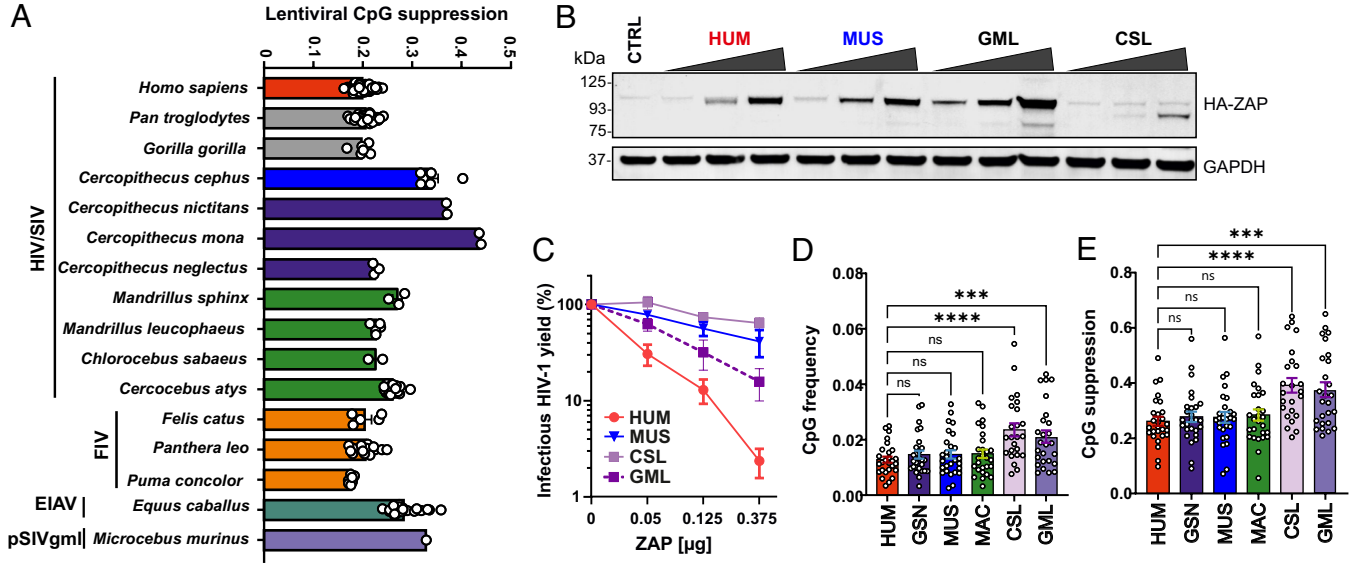

Fig. 6. CpG suppression and ZAP function in lemurs. (*A*) Suppression of CpGs dinucleotides in the genomes of the indicated lentiviruses (*SI Appendix*). (*B*) HEK293T ZAP KO cells were cotransfected with increasing amounts of indicated vectors expressing N-terminally HA-tagged ZAP-L proteins or empty vector control (CTRL) and analyzed by western blot for expression levels. (*C*) HEK293T ZAP KO cells were cotransfected with HIV-1 CH058 proviral construct and increasing amounts of vectors expressing HUM, MUS, or lemur (GML and CSL) ZAP-L isoforms. Infectious virus yields were normalized to the values obtained in the absence of ZAP set as (100%). Data are shown as mean of three to four independent experiments, each tested in triplicates, ± SEM. (*D* and *E*) Comparison of the frequency (*D*) and suppression (*E*) of CpG dinucleotides in 27 innate immune genes from the indicated primate species. Asterisks indicate statistical significance using Dunnett's multiple comparisons test ***$P < 0.001$; ****$P < 0.0001$.

hypothesized that ZAP activity and avoidance of self-targeting by ZAP may also affect CpG suppression in primate host transcripts. Thus, we examined whether the transcripts of GML and CSL are also less CpG suppressed than those of other primate species. We focused on a set of 27 interferon-stimulated genes (ISGs; see *SI Appendix*) as these genes are under increased selection pressure for CpG suppression (52). We found that both lemur species show significantly higher CpG frequencies (Fig. 6*D*) and thus lower suppression (Fig. 6*E*) than humans and other species. Altogether, these data suggest that low antiviral activity of ZAP is associated with reduced host and endogenous lentiviral CpG suppression in lemurs.

Discussion

Suppression of CpG dinucleotides is a hallmark of mammalian genomes and mimicked by numerous viral pathogens. Determining the mechanisms responsible for CpG suppression in viral and cellular genomes is relevant for our understanding of host–pathogen interactions and evolution. Decreases in CpG dinucleotides in vertebrate genomes are thought to result from C-to-T mutations caused by DNA methyltransferases and methyl-cytosine deamination, while CpG suppression in viral genomes may be driven by the evasion of CpG-specific restriction by ZAP. However, there is currently little direct evidence for a role of ZAP in CpG suppression in vivo. In this study, we show that SIVs infecting GSN, mona, and mustached monkeys exhibit a significantly higher CpG content in their genomes compared to other primate lentiviruses, including SIVcpz and SIVgor, which infect chimpanzees and gorillas, respectively. Analysis of an infectious molecular clone of SIV from mustached monkeys revealed that reduced CpG suppression was not due to viral resistance to ZAP inhibition. Rather, mustached and mona monkeys contain a specific mutation in SD4 and express XL forms of ZAP that have poor antiviral activity compared to the L isoform expressed by humans and other primate species. An artificially generated L-form of musZAP also had attenuated antiviral activity, but a combination of five mutations derived from strongly restrictive ZAPs of other primate species fully restored the antiviral activity of musZAP. Finally, we show that an endogenous lentivirus found in gray mouse lemurs has the highest CpG content of all lentiviruses analyzed and this was associated with poor antiviral activity of the corresponding lemur ZAP.

Viral pathogens evolve more rapidly than their mammalian hosts, and the genomes of *Cercopithecus* monkeys, chimpanzees, and gorillas do not differ significantly in their CpG content. However, similar to mustached monkeys, we found that lemurs express ZAPs with poor antiviral activity. Lemurs are only found on Madagascar and nearby smaller islands and diverged from other primate lineages about 50 to 150 Mya (53, 54), while the changes in SD4 promoting expression of ZAP-XL most likely evolved in a common precursor of mustached, mona and other *Cercopithecus* monkeys less than 10 Mya (Fig. 3*D*). Both lemur genomes and their endogenous lentiviruses exhibit less CpG suppression compared to present-day primate lentiviruses and African nonhuman primate species. Since pSIVgml entered the lemur germline several Mya (51), it remains unknown whether it originally displayed higher CpG levels or whether this feature evolved following germline invasion of the lemur genome over evolutionary time due to transcriptional silencing by the action of DNA methyltransferases (55). Nonetheless, it is tempting to speculate that impaired ZAP-dependent degradation of CpG containing RNA constructs also contributes to increased CpG levels of pSIVgml and lemurs.

SIVcpz, the precursor of pandemic HIV-1 group M strains, is a chimeric virus with most of its 3′half genome, including the *env* gene, originated from a precursor of current SIVgsn/mus/mon strains (30). Chimpanzee ZAP-L shows high antiviral activity (Fig. 4*D*) and our data show a decrease in *env* CpG frequency from ~57 CpGs in the SIVmus/mon/gsn clade to ~24 CpGs in SIVcpz (Fig. 1*C*), indicating adaptation after cross-species transmission. SIVcpz was generated by recombination of at least two, and possibly three, different SIVs from smaller monkeys (30, 31), requiring either simultaneous or consecutive infections. While the particular circumstances remain unknown, it seems clear that SIVmus is highly sensitive to ZAP restriction (Fig. 2*E* and *SI Appendix*, Fig. S3). Thus, recombination not only generated a functional *vif* gene (56) but must have also resulted in a significant reduction in ZAP restriction, both of which are required for efficient replication of SIVcpz in its new chimpanzee host.

Our findings suggest that ZAP activity influences the degree of CpG suppression in primate lentiviruses and possibly also their respective host genomes after long evolutionary periods. However, while the number of CpG dinucleotides in viral genomes plays an important role, it is not the only determinant of ZAP sensitivity. The sequence environment and the accessibility of CpG dinucleotides for ZAP binding that is determined by the RNA secondary structure also affects the antiviral effect (57–59). In addition, posttranscriptional methylation of viral RNA may alter recognition of CpG dinucleotides by ZAP. Furthermore, it might be important whether the targeted RNAs are limiting viral infectivity and replication. For example, the CpG frequency in the *env* gene, rather than the overall proviral CpG content, governs the susceptibility of HIV-1 to ZAP (28). Env is translated together with Vpu from a bicistronic RNA, its expression depends on leaky scanning of the *vpu* initiation codon (60), and HIV-1 particles contain only a low number of 7 to 11 Env trimers. Thus, HIV-1 and other *vpu*-encoding lentiviruses might be particularly susceptible to reductions in *env* transcripts due to ZAP-mediated degradation. Notably, some viruses do not evade ZAP restriction by lowering the numbers or accessibility of CpGs but evolved antagonists that counteract ZAP restriction (61–63). We largely focused on ZAP isoforms containing the PARP-like region because previous studies have shown that this domain is critical for CpG-dependent antiviral activity (40, 42, 43). However, we verified that the S isoform of musZAP is also poorly active (*SI Appendix*, Fig. S6). Moreover, the changes in SD4 likely not only favor expression of ZAP-XL over ZAP-L but also M over S. Notably, ZAP XL and M isoforms containing the extended exon 4 generally displayed little antiviral activity arguing against an important role in vivo. Finally, the expression and function of ZAP cofactors, such as KHNYN and TRIM25 (48–50), plays a relevant role in the selection pressures on CpG suppression in viral genomes. While RNA structure and other viral countermeasures may affect sensitivity, our results suggest that impaired antiviral activity of ZAP is associated with higher levels of CpGs in the genomes of primate lentiviruses.

In conclusion, our results support the hypothesis that the antiviral activity of ZAP plays a role in the suppression of CpG dinucleotides in viral and possibly also host genomes in vivo. Since ZAP drives CpG suppression in retroviruses and most likely also retrotransposons, which make up large parts of the human genome, it is likely that ZAP is not only involved in antiviral defenses but also shapes the host genome composition. CpGs play an important role in regulating gene expression and cellular processes (64, 65) and changes in CpG methylation patterns are associated with cancer (66). Thus, future studies on the pros and cons of reduced ZAP activity in primates and their consequences

on CpG frequencies and distribution will provide further insights into the virus–host arms race and the mechanisms that define the composition of our genomes.

Materials and Methods

Lentiviral Sequence Analysis and CpG Mapping. Lentiviral genomes were downloaded from the NIH genetic sequence database (Genbank) and host mRNA sequences were extracted from transcriptomes (NCBI) of primate species (*SI Appendix*). CpG frequencies were calculated as the number of dinucleotides in the viral gene or genome divided by its base pair length. CpG suppression was calculated as detected frequency of CpG dinucleotides divided by the expected frequency of CpG dinucleotides based on the number of C and G residues in the sequence. Numbers of CpGs were calculated using a sliding window of 100 nucleotides and a step size of 100 as reported previously (28). The raw data are displayed as a heat map generated by GraphPad Prism.

Construction of an Infectious Molecular Clone of SIVmus. The consensus sequence for SIVmus-11GabPts2 is available under GenBank accession number KF304708. Among existing SIVmus proviral sequences, this strain was selected for molecular clone construction based of the completeness of its LTR sequence, the presence of only a single ambiguous position across the genome, and for encoding an entry-competent Env protein (67). The proviral sequence was chemically synthesized (GenScript Biotech) as described previously (68). Briefly, the full-length provirus (10,171 bp) was assembled from three non-overlapping subgenomic fragments, ligated using the unique internal BamHI and PacI restriction enzymes. MluI and NotI sites, added to the proviral 5′ and 3′ termini, respectively, enabled the directional cloning into the low-copy-number pBR vector (Fig. 2*B*).

Cell Lines. HEK293T ZAP KO cells have been previously described (11). TZM-bl cells (NIBSC, Cat#ARP5011), which express CD4, CCR5, and CXCR4 and contain the β-galactosidase genes under the control of the HIV promoter(36), have been provided by Drs. Kappes and Wu and Tranzyme Inc. through the NIH AIDS Reagent Program. Cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 2.5% (during virus production) or 10% (at all other times) heat-inactivated fetal calf serum (FCS), 100 U/mL Pen-Strep and 2 mM L-glutamine. Cell lines were regularly tested and confirmed to be negative for *mycoplasma*.

Human PBMCs and CD4+ T Cells. PBMCs from healthy donors were isolated using lymphocyte separation medium (Biocoll separating solution; Biochrom). Negative isolation of CD4+ T cells was carried out using RosetteSep Human CD4+ T Cell Enrichment Cocktail; EasySepTM Human Naïve CD4+ T cell Isolation Kits (Stem Cell Technologies), following the manufacturer's instructions. Primary cells were cultured in RPMI-1640 medium supplemented with 10% FCS, glutamine (2 mM), 100 U/mL Pen-Strep (Gibco), and IL-2 (10 ng/mL) (Miltenyi Biotec). PBMCs and CD4+ T cells were stimulated for 3 d with IL-2 (10 ng/mL) and PHA (2 μg/mL).

Viral Replication Kinetics. One million prestimulated PBMCs and CD4+ T cells were infected with produced virus stocks and normalized for infectivity in TZM-bl cells to the least infectious one. Both unmodified and VSV-G-pseudotyped SIVmus were used for initial infection. Virus stocks were left untreated or incubated with 12.5 μg/mL of fibrils (EF-C) (37) for 10 min at room temperature to boost infection. The following day, cells were washed in PBS to remove residual input virus and the last washing was kept as background control. Samples were transferred to a 96-well plate, virus-containing supernatants were harvested every 3 d, and fresh RPMI 1640 medium was added until the end of culture at day 15. Infectious virus production was measured by TZM-bl cell infectivity assay.

PCR Amplification and Cloning of Primate ZAPs. Peripheral blood mononuclear cells (PBMCs) were isolated from 1 mL of whole blood collected from various primate species. (*SI Appendix,* Table S2), using Biocoll Separating Solution (Biochrom GmbH) and subjected to Biocoll gradient centrifugation, following the manufacturer's protocols. Cells were counted and lysed in RLT Plus buffer containing 1% β-mercaptoethanol, and total RNA was isolated using the RNeasy Plus Mini Kit (Qiagen, Cat #74136) according to the manufacturer's instructions. Residual genomic DNA was removed using the DNA-freeTM DNA Removal Kit (ThermoFisher, Cat #AM1906). RNA concentrations were determined using the NanoDrop 2000 Spectrophotometer and RNAs were subjected to cDNA synthesis using the PrimeScript RT Reagent Kit (Takara, Cat #RR037A) with random 6-mers and oligo dT primers. Reactions without reverse transcriptase were included as controls to exclude contamination with genomic DNA. cDNA was used for ZAP amplification using a mix of the following primer sets: universal primer set 1; Forward-primer#1: CACTTGTAAGTTTCGATTTC; Reverse primer#1: GGCATCCTTCTGACGCTGTATTC, and/or more specific primer set 2: Forward-primer#2: CGATGGCGGACCCGGAGGTGTGC, Reverse primer#2: CTAACTAATCACGCAGGCTTTGTCTG. ZAP PCR products amplified from cDNA of monkeys, apes, and humans were separated using 0.7% (w/v) agarose gel electrophoresis. All primers used in the amplification were obtained from Biomers. The ZAP PCR fragments were ligated into the Zero Blunt™ PCR Cloning Vector (Thermo Fisher, Cat# K2700-20). Plasmid sequencing was performed by Eurofins Genomics or Seqlab.

ZAP Expression and Proviral Constructs. ZAP Plasmids cloned from cDNA were tagged with N-terminal hemagglutinin (HA) tag, and further subcloned into pCG IRES BFP expression plasmid via XbaI/MluI sites. Using the following primer set: forward primer: CTTCTAGAATATGTACCCATACGATGTTCCAGATTACGCTGCGGACCCGGAGGTGTGC and reverse primer: GCTACGCGTCTAACTAATCACGCAGGCTTTGTCTG. The vector encoding human ZAP-L containing N-terminal HA tag was generated by cloning human ZAP (GenScript ORF cDNA clone, Cat #OHu25350) into XbaI/MluI sites of the pCG IRES BFP plasmid, previously described in (28). The sequence of the insert matched the NCBI Reference Sequence: NM_020119.4. Vectors encoding N terminal HA tagged: ZAP-XL from humans and de Brazza's monkey, as well as ZAP-L from gorilla, de Brazza's monkey, Coquerel's sifaka, and gray mouse lemur were synthesized by Twist bioscience and subcloned into the pCG IRES BFP expression vector via XbaI/MluI sites. Sequence analyses confirmed that the ZAP genes match the respective NCBI reference sequences (*SI Appendix,* Table S3). Primate lentiviral infectious molecular clones (IMCs) and their derivatives, along with their corresponding accession numbers, have been previously described (28).

Generation of ZAP M and S Isoforms. Vectors encoding N-terminal HA-tagged ZAP-M (medium) and ZAP-S (short) isoforms from humans, mustached monkeys (*Cercopithecus cephus*), mona monkeys (*Cercopithecus mona*), and De Brazza's monkeys (*Cercopithecus neglectus*) were generated using the Q5 Site-Directed Mutagenesis Kit (NEB, Cat #E0554) following the manufacturer's instructions. ZAP-S constructs were generated from ZAP-L, while ZAP-M constructs were generated from ZAP-XL plasmids. The following primer sets were used for both ZAP-M and ZAP-S: Forward primer (FW): gtaaACGCGTCGGATCCCG, Reverse primers (RV): For Mus, Mon, and Deb monkeys: RV-1 TCTGGCCCTCTCTGCATCTG. For human ZAPs S and M: RV-2 TCTGGCCCTCTCTTCATCTG. All constructs were verified by Sanger sequencing using a commercial sequencing service (Eurofins Genomics or Seqlab).

Assessing ZAP Isoforms-Mediated Viral Restriction. To assess the relative sensitivity of HIV-1 CH058 to restriction by the four ZAP isoforms–extralong (XL), long (L), medium (M), and short (S)–from different species, we tested ZAPs derived from humans, mustached monkeys, mona monkeys, and De Brazza's monkeys. HEK293T ZAP KO cells (seeded in 24-well format) were cotransfected using PEI transfection reagent with 625 ng of the proviral HIV-1 and indicated increasing concentration of ZAP expression vectors as indicated in each figure. The total amount of DNA was normalized to 1 μg by adding empty pCG IRES BFP vector. Culture supernatants were harvested 2 d later and used to infect TZM-bl cells. β-galactosidase activity was measured *t* days later using the Gal-Screen kit (Applied Biosystems) as relative light units per second using microplate luminometer. Infectious virus yield values of each IMC in the presence of ZAP were normalized to the corresponding BFP-only control (100%).

Site-Directed Mutagenesis. pCG_humZAP-L HA IRES BFP, pCG_musZAP-L HA IRES BFP and pCG_musZAP-XL HA IRES BFP mutant plasmids were generated using the Q5 Site-directed Mutagenesis Kit (NEB, Cat #E0554). Nucleotide changes introducing single amino acid change between fully active ZAPs and attenuated musZAPs were introduced into mus ZAPs-L and -XL and vice versa into humZAP-L isoform (see *SI Appendix,* Table S4 for primer sequences). Constructs encoding combined five amino acid changes mutants were synthesized by Twist bioscience and subcloned into the pCG IRES BFP expression vector via XbaI/MluI sites. All constructs were verified by sequence analysis using a commercial sequencing service (Eurofins Genomics or Seqlab).

Viral Sensitivity to ZAP Restriction. To assess relative lentiviral sensitivity to ZAP, HEK293T ZAP KO cells (seeded in 24-well format) were cotransfected using PEI transfection reagent with 625 ng of the proviral HIV-1 and SIV constructs and increasing concentration of ZAP expression vectors as indicated in each figure. The total amount of DNA was normalized to 1 μg by adding empty pCG IRES BFP vector. Culture supernatants were harvested 2 d later and used to infect TZM-bl cells. β-galactosidase activity was measured *t* days later using the Gal-Screen kit (Applied Biosystems) as relative light units per second using microplate luminometer. Infectious virus yield values of each IMC in the presence of ZAP were normalized to the corresponding BFP-only control (100%).

Prediction of Splice Donor Strength. Publicly available whole genome sequencing (WGS) data and genomes were used to identify genomic regions containing ZAP coding sequences in different monkey species. Human ZAP (genomic sequence from hg38 assembly (chr7:139,043,515–139,109,720) available at the UCSC genome browser; https://genome.ucsc.edu/) was used as a bait for the BLASTn searches. Splice donor and acceptor usage was predicted with SpliceRover (http://bioit2.irc.ugent.be/rover/splicerover/) (44), using default settings for the human models.

Western Blots. To examine viral protein expression levels under the expression of ZAPs, HEK293T ZAP KO cells were cotransfected with the indicated HIV-1 proviral constructs and pCG HA-ZAP IRES BFP DNA vector or empty pCG IRES BFP vector control. 2 d posttransfection, cells were lysed with Co-IP buffer (150 mM NaCl, 50 mM HEPES, 5 mM EDTA, 0.10% NP40, 0.5 mM sodium orthovanadate, 0.5 mM NaF, protease inhibitor cocktail from Roche) and cell-free virions were purified by centrifugation of cell culture supernatants through a 20% sucrose cushion at 20,800×*g* for 90 min at 4 °C and lysed in Co-IP lysis buffer. Samples were reduced in the presence of β-mercaptoethanol by boiling at 95 °C for 10 min. Proteins were separated in 4 to 12% Bis-Tris gradient acrylamide gels (Invitrogen), blotted onto polyvinylidene difluoride (PVDF) membrane, and incubated with anti-HIV-1 Env (obtained through the NIH AIDS Reagent Program, Cat #12559), anti-p24 (Abcam, Cat #ab9071), anti-HA tag (Cell Signaling, Cat #C29F4), anti-GAPDH (Santa Cruz, Cat #sc-365062), and anti-ZAP (GeneTex, Cat #GTX120134) antibodies. Subsequently, blots were probed with IRDye® 680RD Goat anti-Rabbit IgG (H + L) (Cat #926-68071; LI-COR), IRDye® 800CW Goat anti-Mouse IgG (H + L) (LI-COR, Cat #926-32210) and IRDye 800RD Goat anti-Rat IgG (H + L) (LI-COR, Cat #925-32219) Odyssey antibodies and scanned using a LI-COR Odyssey reader.

Innate Immune Genes Selected for CpG Analysis. CpG mapping was conducted in 27 genes, 23 of them were previously identified as CpG-suppressed interferon-stimulated genes (ISGs) (52), the selected genes were CXCL10, CCL8, IFI44L, CXCL11, SAMD9L, BATF2, IFIT1, TNFSF10, RTP4, IDO1, IFIT3, OASL, OAS2, TNFSF13B, DDX58, MX2, TLR3, RSAD2, OAS1, CXCL9, C3AR1, IFIT2, and AIM2. In addition, we included CD4, ZC3HAV1, KHNYN, and TRIM25 in the panel. *M. mulatta*, *Homo sapiens*, *P. coquereli,* and *M. murinus* transcript sequences were acquired from the NCBI database (https://www.ncbi.nlm.nih.gov/nuccore). The genome of *Cercopithecus nictitans* was fully sequenced by the group of Prof. Steven. E Bosinger at the Emory National Primate Research at Emory University in Atlanta. The genome of rhesus macaques was used as a reference to complete the full-length sequencing of GSN by next-generation sequencing (NGS). Data were analyzed using IGV_2.16.1 software. The MUS transcript sequences of interest were identified by BLASTn searches of whole genome sequencing (WGS) data deposited at the NCBI using sequences *M. mulatta* transcripts as a bait.

ZAPs Amino Acid Alignments and Sequence Diversity. Primate ZC3HAV1 (ZAP) sequences reported in this study or previously published in the NCBI sequence database (NCBI orthologs) were aligned to human ZAP-L or ZAP XL as indicated in *SI Appendix*, Figs. S5, S6, and S9 using Multalin tool (http://multalin.toulouse.inra.fr/multalin/multalin.html). Amino acids at the positions corresponding to positions V120, N123, A306, I457, and T605 in humZAP-L were extracted and grouped according to the evolutionary relationship of the corresponding primate species.

Co-IP. For coCo-IP, ZAP-HA transfected cells were washed, harvested, and pooled by detaching cells using 0.05% Trypsin/EDTA. Cell pellets were resuspended in 1 mL Co-IP buffer supplemented with ULTRA protease inhibitor cocktail tablet and incubated on ice for 20 min. Next, cells were centrifuged (4,000 rpm, 4 °C, 20 min), supernatants (SN) pooled and 60 μL kept as input control. Next, 2.5 μL of appropriate primary antibody was added to the Co-IP samples and incubated at 4 °C. The antigen/antibody mixture was added to washed Pierce Protein A/G Magnetic Beads (Thermo Fisher Scientific, Cat #88802) and incubated at RT for 4 h, rolling. Then, samples were placed on the magnetic stand, supernatant discarded, and samples washed 3x with 1 mL NP40 wash buffer. The last washing step included 500 μL sterile water. Next, 60 μL 1x Loading Dye diluted in Co-IP buffer was added and samples were boiled at 95 °C for 10 min. Co-IP samples were once again placed onto the magnetic stand, loading dye was collected and 1.75 μL β-mercaptoethanol was added. Co-IP data were analyzed using western blot analysis by pulling down ZAP-HA tag and detecting endogenous KHNYN (anti-KHNYN, Santa Cruz Biotechnology, Cat #sc-514168), as well as TRIM25 (anti-TRIM25, BD Biosciences, Cat #610570).

Statistical Analysis. Statistical analyses were performed using GraphPad Prism software. Two-tailed unpaired Student's *t* test or Mann–Whitney U-test was used to determine statistical significance. Spearman's nonparametric test was used for correlation analyses. Dunnett's multiple comparisons test was used for CpG frequency and suppression analyses between different species. Significant differences are indicated as $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$. Statistical parameters are specified in the figure legends.

Ethics Statement

Our research complies with all relevant ethical regulations. Experiments involving the isolation and differentiation of PBMCs and CD4+ T cells from peripheral blood were approved by the ethics committee of Ulm University (approval Cat #151/22). All donors were anonymous and provided written informed consent. Monkey blood was collected during routine health examinations or from animals that died of natural causes.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*. All raw data associated with this study, including CpG calculations, Western blot membranes, electrophoresis gels, and ZAP sequences have been deposited in Figshare and are publicly available at: https://figshare.com/articles/dataset/Raw_Data_for_Host_ZAP_Activity_Correlates_with_the_Levels_of_CpG_Suppression_in_Primate_Lentiviruses/28603472 (69).

ACKNOWLEDGMENTS. We thank Drs. Baptiste Mulot and Antoine Leclerc (Zoo Parc de Beauval, Saint-Aignan, France) for providing monkey blood. F.K. received support from an Advanced European Research Council grant (Traitor-viruses). K.M.J.S. acknowledges funding by the German Federal Ministry of Education and Research (BMBF; IMMUNOMOD-01KI2014). F.K. and K.M.J.S. receive funding by the German Research Foundation (DFG) via Project ID 316249678 (SFB 1279) and KI 548/21-1 (F.K.) as well as SP 1600/7-1/9-1 (K.M.J.S). D.S. was funded by the Heisenberg Program (grant ID: SA 2676/3-1) of the DFG. B.H.H, is funded by the NIH (R01 AI 162646, UM1 AI 164570). Sequences for *C. nictitans* ZAP were provided by the Emory Non-human primate research center Genomics Core which is supported in part by NIH P51 OD011132 using an Illumina NovaSeq 6000 funded by NIH S10 OD026799 (S.E.B.). D.K. received funding from Marie Sklodowska-Curie European Union's Horizon 2020 program (grant agreement No. 101062524) and the Else-Krönung Fresenius Stiftung (2022_EKEA.47). R.N. and D.K. acknowledge funding by the "Baustein" program of Ulm University Medical Faculty.

Author affiliations: [a]Institute of Molecular Virology, Ulm University Medical Center, Ulm 89081, Germany; [b]Research group "Mechanisms of innate Antiviral immunity", Institute for Medical Virology and Epidemiology of Viral Diseases, University Hospital Tübingen, Tübingen 72076, Germany; [c]Department of Medicine and Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; [d]Neurovirology & Neuroinflammation, German Center for Neurodegenerative Diseases (DZNE), Ulm 89081, Germany; [e]Zoo La Palmyre, Les Mathes 17570, France; [f]Zoo Sanary-sur-Mer, Sanary-sur-Mer 83110, France; and [g]Department of Pathology & Laboratory Medicine, Emory University, Division of Microbiology and Immunology, Emory National Primate Research Center, Atlanta, GA 30329

1. M. Sironi, R. Cagliani, D. Forni, M. Clerici, Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nat. Rev. Genet.* **16**, 224–236 (2015).
2. M. Schlee, G. Hartmann, Discriminating self from non-self in nucleic acid sensing. *Nat. Rev. Immunol.* **16**, 566–580 (2016).
3. H.-C. Lee, K. Chathuranga, J.-S. Lee, Intracellular sensing of viral genomes and viral evasion. *Exp. Mol. Med.* **51**, 1–13 (2019).
4. N. K. Duggal, M. Emerman, Evolutionary conflicts between viruses and restriction factors shape immunity. *Nat. Rev. Immunol.* **12**, 687–695 (2012).
5. F. P. Lobo *et al.*, Virus-host coevolution: Common patterns of nucleotide motif usage in flaviviridae and their hosts. *PLoS ONE* **4**, e6282 (2009).
6. B. D. Greenbaum, A. J. Levine, G. Bhanot, R. Rabadan, Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLoS Pathog.* **4**, e1000079 (2008).
7. L. P. Villarreal, G. Witzany, Viruses are essential agents within the roots and stem of the tree of life. *J. Theor Biol.* **262**, 698–710 (2010).
8. O. G. Pybus, A. Rambaut, Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* **10**, 540–550 (2009).
9. S. Karlin, J. Mrázek, Compositional differences within and between eukaryotic genomes. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10227–10232 (1997).
10. S. Karlin, W. Doerfler, L. R. Cardon, Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *J. Virol.* **68**, 2889–97 (1994).
11. M. A. Takata *et al.*, CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* **550**, 124–127 (2017).
12. G. Gao, X. Guo, S. P. Goff, Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* **297**, 1703–1706 (2002).
13. X. Guo, J.-W.N. Carroll, M. R. Macdonald, S. P. Goff, G. Gao, The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *J. Virol.* **78**, 12781–12787 (2004).
14. M. J. Bick *et al.*, Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J. Virol.* **77**, 11555–11562 (2003).
15. R. Mao *et al.*, Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. *PLoS Pathog.* **9**, e1003494 (2013).
16. Q. Tang, X. Wang, G. Gao, The short form of the zinc finger antiviral protein inhibits influenza A virus protein expression and is antagonized by the virus-encoded NS1. *J. Virol.* **91**, e01909–16 (2017).
17. S. Müller *et al.*, Inhibition of filovirus replication by the zinc finger antiviral protein. *J. Virol.* **81**, 2391–2400 (2007).
18. J. Schwerk *et al.*, RNA-binding protein isoforms ZAP-S and ZAP-L have distinct antiviral and immune resolution functions. *Nat. Immunol.* **20**, 1610–1620 (2019).
19. S. Hayakawa *et al.*, ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. *Nat. Immunol.* **12**, 37–44 (2011).
20. E. S. Lander *et al.*, Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
21. J. C. Venter *et al.*, The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
22. D. N. Cooper, M. Krawczak, Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum. Genet.* **83**, 181–188 (1989).
23. P. A. Jones, S. B. Baylin, The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**, 415–428 (2002).
24. D. Gonçalves-Carneiro, M. A. Takata, H. Ong, A. Shilton, P. D. Bieniasz, Origin and evolution of the zinc finger antiviral protein. *PLoS Pathog.* **17**, e1009545 (2021).
25. P. M. Sharp, B. H. Hahn, Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspect. Med.* **1**, a006841 (2011).
26. I. Pandrea, D. L. Sodora, G. Silvestri, C. Apetrei, Into the wild: Simian immunodeficiency virus (SIV) infection in natural hosts. *Trends Immunol.* **29**, 419–428 (2008).
27. D. Sauter, F. Kirchhoff, Key viral adaptations preceding the AIDS Pandemic. *Cell Host Microbe* **25**, 27–38 (2019).
28. D. Kmiec *et al.*, CpG frequency in the 5′ third of the env gene determines sensitivity of primary HIV-1 strains to the zinc-finger antiviral protein. *mBio* **11**, e02903 (2020).
29. F. Gao *et al.*, Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature* **397**, 436–441 (1999).
30. E. Bailes *et al.*, Hybrid origin of SIV in chimpanzees. *Science* **300**, 1713 (2003).
31. S. M. Bell, T. Bedford, Modern-day SIV viral diversity generated by extensive recombination and cross-species transmission. *PLoS Pathog.* **13**, e1006466 (2017).
32. F. Van Heuverswyn *et al.*, Human immunodeficiency viruses: SIV infection in wild gorillas. *Nature* **444**, 164 (2006).
33. C. Gilbert, D. G. Maxfield, S. M. Goodman, C. Feschotte, Parallel germline infiltration of a lentivirus in two malagasy lemurs. *PLOS Genet.* **5**, e1000425 (2009).
34. V. Courgnaud *et al.*, Identification of a new simian immunodeficiency virus lineage with a vpu gene present among different cercopithecus monkeys (C. mona, C. cephus, and C. nictitans) from Cameroon. *J. Virol.* **77**, 12523–12534 (2003).
35. F. Liégeois *et al.*, Full-length genome analyses of two new simian immunodeficiency virus (SIV) strains from moustached monkeys (*C. Cephus*) in gabon illustrate a complex evolutionary history among the SIVmus/mon/gsn lineage. *Viruses* **6**, 2880–2898 (2014).
36. E. J. Platt, K. Wehrly, S. E. Kuhmann, B. Chesebro, D. Kabat, Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. *J. Virol.* **72**, 2855–2864 (1998).
37. M. Yolamanova *et al.*, Peptide nanofibrils boost retroviral gene transfer and provide a rapid means for concentrating viruses. *Nat. Nanotechnol.* **8**, 130–136 (2013).
38. F. V. Heuverswyn *et al.*, Genetic diversity and phylogeographic clustering of SIVcpzPtt in wild chimpanzees in Cameroon. *Virology* **368**, 155–171 (2007).
39. M. M. H. Li *et al.*, Characterization of novel splice variants of zinc finger antiviral protein (ZAP). *J. Virol.* **93**, e00715 (2019), 10.1128/JVI.00715-19.
40. J. A. Kerns, M. Emerman, H. S. Malik, Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genet.* **4**, e21 (2008).
41. G. Charron, M. M. H. Li, M. R. MacDonald, H. C. Hang, Prenylome profiling reveals S-farnesylation is crucial for membrane targeting and antiviral activity of ZAP long-isoform. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11085–11090 (2013).
42. D. Kmiec, M. J. Lista, M. Ficarelli, C. M. Swanson, S. J. D. Neil, S-farnesylation is essential for antiviral activity of the long ZAP isoform against RNA viruses with diverse replication strategies. *PLoS Pathog.* **17**, e1009726 (2021).
43. G. Xue *et al.*, Poly(ADP-ribose) potentiates ZAP antiviral activity. *PLoS Pathog.* **18**, e1009202 (2022).
44. J. Zuallaert *et al.*, SpliceRover: Interpretable convolutional neural networks for improved splice site prediction. *Bioinformatics* **34**, 4180–4188 (2018).
45. Y. Shao *et al.*, Phylogenomic analyses provide insights into primate evolution. *Science* **380**, 913–924 (2023).
46. C. Ochsenbauer *et al.*, Generation of transmitted/founder HIV-1 infectious molecular clones and characterization of their replication capacity in CD4 T lymphocytes and monocyte-derived macrophages. *J. Virol.* **86**, 2715–2728 (2012).
47. S. Huang *et al.*, Positive selection analyses identify a single WWE domain residue that shapes ZAP into a more potent restriction factor against alphaviruses. *PLOS Pathog.* **20**, e1011836 (2024).
48. M. M. H. Li *et al.*, TRIM25 enhances the antiviral action of zinc-finger antiviral protein (ZAP). *PLoS Pathog.* **13**, e1006145 (2017).
49. X. Zheng *et al.*, TRIM25 is required for the antiviral activity of zinc finger antiviral protein. *J. Virol.* **91**, e00088 (2017).
50. M. Ficarelli *et al.*, KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *eLife* **8**, e46767 (2019).
51. R. J. Gifford *et al.*, A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20362–20367 (2008).
52. A. E. Shaw *et al.*, The antiviral state has shaped the CpG composition of the vertebrate interferome to avoid self-targeting. *PLOS Biol.* **19**, e3001352 (2021).
53. A. D. Yoder, M. Cartmill, M. Ruvolo, K. Smith, R. Vilgalys, Ancient single origin for Malagasy primates. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5122–5126 (1996).
54. J. Shoshani, C. P. Groves, E. L. Simons, G. F. Gunnell, Primate phylogeny: Morphological vs. molecular results. *Mol. Phylogenet. Evol.* **5**, 102–154 (1996).
55. W. Zhou, G. Liang, P. L. Molloy, P. A. Jones, DNA methylation enables transposable element-driven genome expansion. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 19359–19366 (2020).
56. L. Etienne, B. H. Hahn, P. M. Sharp, F. A. Matsen, M. Emerman, Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. *Cell Host Microbe* **14**, 85–92 (2013).
57. J. L. Meagher *et al.*, Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 24303–24309 (2019).
58. S. Chen *et al.*, Structure of N-terminal domain of ZAP indicates how a zinc-finger protein recognizes complex RNA. *Nat. Struct. Mol. Biol.* **19**, 430–435 (2012).
59. X. Luo *et al.*, Molecular mechanism of RNA recognition by zinc-finger antiviral protein. *Cell Rep.* **30**, 46–52.e4 (2020).
60. S. Schwartz, B. K. Felber, E. M. Fenyö, G. N. Pavlakis, Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. *J. Virol.* **64**, 5448–5456 (1990).
61. C. Su, J. Zhang, C. Zheng, Herpes simplex virus 1 UL41 protein abrogates the antiviral activity of hZAP by degrading its mRNA. *Virol. J.* **12**, 203 (2015).
62. L. Xie *et al.*, The 3C protease of enterovirus A71 counteracts the activity of host zinc-finger antiviral protein (ZAP). *J. Gen Virol.* **99**, 73–85 (2018).
63. Y. Xuan *et al.*, ZAP inhibits murine gammaherpesvirus 68 ORF64 expression and is antagonized by RTA. *J. Virol.* **87**, 2735–2743 (2013).
64. L. D. Moore, T. Le, G. Fan, DNA methylation and its basic function. *Neuropsychopharmacol* **38**, 23–38 (2013).
65. J. A. Cain, B. Montibus, R. J. Oakey, Intragenic CpG Islands and their impact on gene regulation. *Front Cell Dev. Biol.* **10**, 832348 (2022).
66. A. Nishiyama, M. Nakanishi, Navigating the DNA methylation landscape of cancer. *Trends Genet.* **37**, 1012–1027 (2021).
67. R. M. Russell *et al.*, CD4 receptor diversity represents an ancient protection mechanism against primate lentiviruses. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025914118 (2021).
68. H. J. Barbian *et al.*, Neutralization properties of simian immunodeficiency viruses infecting chimpanzees and gorillas. *mBio* **6**, e00296–15 (2015).
69. R. Nchioua, F. Kirchhoff, Data from "Host ZAP activity correlates with the levels of CpG suppression in primate lentiviruses." Figshare. Available at https://figshare.com/articles/dataset/Raw_Data_for_Host_ZAP_Activity_Correlates_with_the_Levels_of_CpG_Suppression_in_Primate_Lentiviruses/28603472. Deposited 16 March 2025.