



Positive selection in dNTPase SAMHD1 throughout mammalian evolution

Christopher Monit^a, Elizabeth R. Morris^b, Christopher Ruis^a, Bart Szafran^c, Grant Thiltgen^a, Ming-Han Chloe Tsai^d, N. Avrion Mitchison^a, Kate N. Bishop^d, Jonathan P. Stoye^c, Ian A. Taylor^b, Ariberto Fassati^{a,1}, and Richard A. Goldstein^{a,1}

^aDivision of Infection and Immunity, University College London, WC1E 6BT London, United Kingdom; ^bMacromolecular Structure Laboratory, The Francis Crick Institute, NW1 1AT London, United Kingdom; ^cRetrovirus–Host Interactions Laboratory, The Francis Crick Institute, NW1 1AT London, United Kingdom; and ^dRetroviral Replication Laboratory, The Francis Crick Institute, NW1 1AT London, United Kingdom

Edited by John M. Coffin, Tufts University, Boston, MA, and approved August 5, 2019 (received for review May 23, 2019)

The vertebrate protein SAMHD1 is highly unusual in having roles in cellular metabolic regulation, antiviral restriction, and regulation of innate immunity. Its deoxynucleoside triphosphohydrolase activity regulates cellular dNTP concentration, reducing levels below those required by lentiviruses and other viruses to replicate. To counter this threat, some primate lentiviruses encode accessory proteins that bind SAMHD1 and induce its degradation; in turn, positive diversifying selection has been observed in regions bound by these lentiviral proteins, suggesting that primate SAMHD1 has coevolved to evade these countermeasures. Moreover, deleterious polymorphisms in human SAMHD1 are associated with autoimmune disease linked to uncontrolled DNA synthesis of endogenous retroelements. Little is known about how evolutionary pressures affect these different SAMHD1 functions. Here, we examine the deeper history of these interactions by testing whether evolutionary signatures in SAMHD1 extend to other mammalian groups and exploring the molecular basis of this coevolution. Using codon-based likelihood models, we find positive selection in SAMHD1 within each mammal lineage for which sequence data are available. We observe positive selection at sites clustered around T592, a residue that is phosphorylated to regulate SAMHD1 activity. We verify experimentally that mutations within this cluster affect catalytic rate and lentiviral restriction, suggesting that virus–host coevolution has required adaptations of enzymatic function. Thus, persistent positive selection may have involved the adaptation of SAMHD1 regulation to balance antiviral, metabolic, and innate immunity functions.

SAMHD1 | HIV-1 | restriction | evolution | mammals

The parasitic nature of their lifestyle brings viruses into evolutionary conflict with the immune systems of their hosts. Vertebrates have evolved an arsenal of innate immunity proteins, called restriction factors, that target conserved features of virus replication cycles, while some viruses, in turn, have evolved means of neutralizing (or “antagonizing”) them, often by mechanisms involving direct protein–protein interactions (1, 2). This leads to an evolutionary “arms race” as the restriction factor undergoes rapid evolution to alter the interaction interface and prevent recognition by a viral antagonist, while the antagonist similarly evolves to restore binding.

SAMHD1 (sterile alpha motif and histidine-aspartic acid domain-containing protein 1) is a restriction factor of several groups of retroviruses and DNA viruses, including lentiviruses [namely, HIV, simian immunodeficiency virus (SIV), and feline immunodeficiency virus (FIV)], vaccinia, herpes simplex 1, and hepatitis B viruses (3–10). Its deoxynucleoside triphosphohydrolase (dNTP-tpase) activity suppresses viral replication by hydrolyzing dNTPs, reducing the intracellular concentration of substrates required for viral DNA production (11, 12). HIV-2 and related SIVs counter SAMHD1 by expressing the accessory protein Vpx that recruits SAMHD1 to DCAF1, targeting it for degradation through the cellular Cullin-4-based E3 ubiquitin ligase machinery (3, 4, 13–16). Some other primate lentiviruses use the related Vpr protein to

fulfill the same role (17), although HIV-1 Vpr does not have the equivalent function. Vpx/Vpr from different lentivirus lineages target different regions of SAMHD1, recognizing either the N or C termini (18). Evolutionary analyses of primate SAMHD1 have shown that positive diversifying selection has occurred in these 2 different binding regions, suggesting an evolutionary arms race between viruses and SAMHD1 in primates (17, 19). SAMHD1 antagonism by primate lentiviruses is often strikingly host-specific, including adaptation to dominant SAMHD1 alleles within species, suggesting that the evolutionary conflict has led to highly intricate coevolution (20).

In addition to its antiviral function, SAMHD1 also maintains the fine balance of intracellular dNTP levels that allows progression of the cell cycle (21), while preventing the accumulation of endogenous nucleic acids (22). The enzyme’s activity is regulated by conversion between the catalytically active tetrameric state and the weakly active monomeric or dimeric forms (23). Tetramers are favored in the presence of SAMHD1’s allosteric regulators, dNTP and GTP/dGTP molecules (24, 25), while phosphorylation of threonine residue 592 (T592), located near the C terminus, reduces the stability of the SAMHD1 tetramer, favoring the monomeric state. In both primates and mice, phosphorylation is mediated by CDKs 1/2 complexed with cyclin

Significance

Animals defend themselves from viral infection using innate immunity proteins that disrupt various stages of the virus life cycle. In response, viruses produce proteins that bind these host factors and compromise their activity, resulting in evolutionary conflict as immunity and virus proteins adapt to prevent and restore binding, respectively. We report that evolutionary conflict involving the host innate immunity protein SAMHD1 has occurred throughout mammalian evolution. We observe adaptation in a region of SAMHD1 that regulates its activity, and we demonstrate how mutations here influence its enzymatic properties, suggesting that evolutionary conflict has involved modulation of SAMHD1 regulation and function. This correlates with reduced restriction of HIV-1, indicating that positive selection has influenced both SAMHD1’s dNTPase and antiviral activities.

Author contributions: C.M., N.A.M., K.N.B., J.P.S., I.A.T., A.F., and R.A.G. designed research; C.M., E.R.M., C.R., B.S., M.-H.C.T., K.N.B., and I.A.T. performed research; C.M., E.R.M., C.R., B.S., G.T., K.N.B., J.P.S., I.A.T., A.F., and R.A.G. analyzed data; and C.M., A.F., and R.A.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: a.fassati@ucl.ac.uk or r.goldstein@ucl.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1908755116/-DCSupplemental.

A2, suggesting that this mechanism of regulation is conserved among mammals (26–30).

Two crucial features of this molecular arms race remain unclear. First, since SAMHD1 is found throughout vertebrates, and DNA-producing viruses infect all domains of life, how widespread is the evolutionary conflict between viruses and SAMHD1 in other taxa? Second, how has SAMHD1 responded to selective pressure from its dual roles in virus restriction and dNTP regulation?

To address these questions, we applied codon-based likelihood models to a large set of SAMHD1 sequences from a diverse range of mammals. We found evidence of positive diversifying selection in every group of mammals for which data are available, indicating a pathogen–SAMHD1 arms race extending throughout mammalian evolution. Strikingly, many of the sites under positive selection cluster around T592, indicating positive selection acting on sites that modulate SAMHD1 phosphorylation, tetramerization, and, therefore, enzymatic activation. We show that replacing amino acids at some of these sites with residues observed in other mammal species reduces dNTP-tpase activity and can reduce HIV-1 restriction in cell culture. SAMHD1 has therefore experienced an unusual combination of selective constraints as selection pressure imposed by viruses interacted with the need to maintain, regulate, and adjust enzymatic activity.

Results

Positive Selection in Mammals. To investigate the history of SAMHD1 during mammalian evolution, we compiled a dataset of 120 publicly available mammalian SAMHD1 coding sequences (*SI Appendix*,

Table S1), including 5 well-represented clades: the Primates ($n = 55$), the Glires (rodents, rabbits, and hares; $n = 16$), the Cetartiodactyla (whales and even-toed ungulates; $n = 18$), the Carnivora (cats, dogs, bears, etc.; $n = 8$), and the Chiroptera (bats; $n = 6$). A phylogenetic tree estimated from these gene sequences by maximum likelihood was mostly concordant with the reported mammalian species phylogeny (31), and the majority of nodes had support values above 70% (*SI Appendix*, Fig. S1).

Using the site-specific selection models implemented in PAML (32–34), we found that the likelihood-ratio test supported the presence of positive selection in mammalian SAMHD1 ($P = 4 \times 10^{-90}$; *SI Appendix*, Table S2), and 36 sites were identified as under positive selection (posterior probability > 0.95; Fig. 1*A* and *SI Appendix*, Table S3). We identified most of the same set of sites as under positive selection when repeating the analysis using 3 alternative tree topologies (*SI Appendix*, Tables S2 and S4), indicating that the result is not sensitive to possible minor inaccuracies in the estimated phylogeny topology.

As positive selection has been detected in primate SAMHD1 (17, 19), we sought to test whether this signature of a molecular arms race was specific to primates by repeating the analysis with the primate clade removed. Again, we found statistical support for positive selection ($P = 3 \times 10^{-66}$), and 31 sites were identified as under positive selection in nonprimate mammals (Fig. 1*B* and *SI Appendix*, Table S3). Of these, 29 had also been identified in the analysis of the all-mammals dataset (*SI Appendix*, Fig. S2*A*). These results indicate that positive selection is not confined to primate SAMHD1, but has also occurred in other mammals.

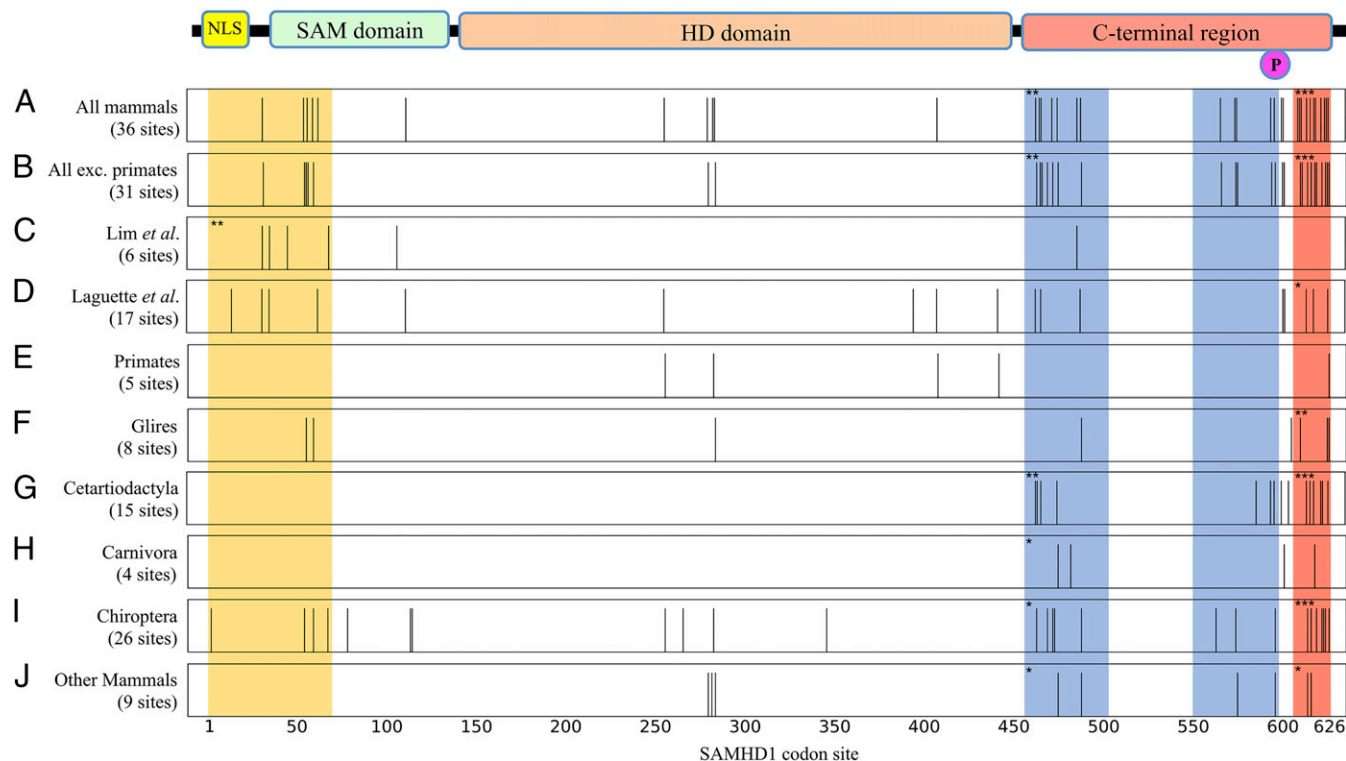


Fig. 1. Codon sites under positive selection in mammalian SAMHD1 identified in this and previous site studies. Panels represent the linear sequence, and vertical bars mark sites where the posterior probability of positive selection (Bayes empirical Bayes calculation) is greater than 0.95 (bar height is not meaningful). In parentheses are the total numbers of sites thus identified. Site numbering is based on the human sequence. At the top is a cartoon showing SAMHD1 sequence features. NLS, nuclear localization signal; P, phosphorylation site T592. Shaded regions correspond to the N-terminal Vpx/Vpr binding site (yellow), C-terminal Vpx/Vpr binding site (red), and region around the T592 phosphorylation site (blue), determined with reference to SAMHD1 crystal structures. Asterisks indicate where there is statistically significant clustering of sites identified as under positive selection in these sequence regions (*SI Appendix*, Table S5). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Note that the 2 sections comprising the phosphorylation region (blue) are treated as a single region. (A) Results from analysis of all mammals with PAML model M2a. (B) Similar analysis of mammals with primates excluded. (C and D) Published results of Lim et al. (17) and Laguette et al. (19), both from analyses of primate SAMHD1 with PAML model M8. (E–J) Sites identified in mammal subgroups, model M8.

Positive Selection in Mammal Subgroups. We next determined whether the positive-selection signal was due to specific groups within mammals by repeating the analyses with each of the well-represented subclades. We also performed the analysis on a 6th group comprising the other species not belonging to a well-represented monophyletic set ($n = 17$), hereafter called “Other Mammals.” We observed evidence for positive selection in all of these subgroups. The identified sites for each subgroup were distributed in similar regions as sites found in our analysis of all mammals (Fig. 1 and *SI Appendix, Table S3*), although many of the identified sites were specific to a particular clade (*SI Appendix, Fig. S2B*). The Chiroptera (bats) had the greatest preponderance of sites, with 26 identified, while the Carnivora (dogs, cats, etc.) had the fewest sites, 4, all in the C-terminal region. These data indicate that positive selection has occurred in multiple groups throughout mammalian evolution. Only 5 sites were identified in the Primates (Fig. 1E), 4 of which were found by Laguette et al. (19) in their analyses of primate SAMHD1 using the same PAML models, while none of these were identified by Lim et al. (17) in a similar analysis, both presumably because of the larger number of sequences now available to us (35). Several of the sites reported by these authors as having posterior probabilities for positive selection above the conventional 0.95 threshold nonetheless had probabilities >0.90 in our analysis (*SI Appendix, Fig. S3*).

Positive Selection at Vpx/Vpr Binding Regions. We next examined whether the sites under positive selection in mammalian SAMHD1 are in the 2 distinct regions of the protein targeted by known SAMHD1 antagonists. We therefore mapped the identified sites onto crystal structures of primate SAMHD1 bound to lentiviral Vpx/Vpr proteins. We found 11 sites under positive selection in all mammals within the C-terminal Vpx/Vpr binding region (sites 606–626; ref. 36; *SI Appendix, Fig. S4A*), 5 of which have side chains directly contacting Vpx residues (sites 609; 610; 611; 618; and 622). By calculating the probability that 11 of 36 identified sites should fall in this binding region comprising 20 sites of a total 626 sites by chance (*Methods*), we found the concentration of sites identified as under positive selection in this region to be highly significant ($P < 10^{-6}$). All but 1 of these Vpx-contacting sites (site 609) were also identified when primate sequences were excluded from the analysis, and the clustering of sites under positive selection in this region (calculated by the same approach) was also significant for sites identified in Glires, Cetartiodactyla, Chiroptera, and Other Mammals alone. We also identified several sites (32; 55; 57; 60; and 63) under positive selection across mammals in the N-terminal Vpx/Vpr binding region (sites 1–69; ref. 37; *SI Appendix, Fig. S4B*), where positive selection was similarly identified by Laguette et al. (19) and Lim et al. (17). This included sites in direct contact with Vpx, though this clustering was not statistically significant. Again, all but 1 of these (site 63) were also identified with primate sequences excluded. This overlap of sites under positive selection in mammals with regions bound by lentiviral Vpx/Vpr proteins, particularly the C-terminal binding site, suggests the existence of factors expressed by other viruses that may target similar regions of SAMHD1 in species other than primates.

Positive Selection in C-Terminal Region Around Phosphorylation Site T592. We observed clustering of sites under positive selection in 2 stretches of the C-terminal region (sites 456–502 and 550–599) that fold together to form a domain containing the phosphorylation site, T592 (Figs. 1 and 2). This clustering was statistically significant ($P < 0.05$) in all mammals, the Cetartiodactyla (whales and even-toed ungulates), the Chiroptera (bats), and the Other Mammals (*SI Appendix, Table S5*). None of the identified sites were located at the interface between SAMHD1 monomers or at the dNTP-binding catalytic or allosteric sites (Fig. 2A). The

positioning of sites under positive selection in the region around the phosphorylation site therefore suggests that rapid evolution involved modulation of SAMHD1 function.

Point Mutations at C-Terminal Sites under Positive Selection Affect SAMHD1 Function. Given the interesting distribution of sites under positive selection, we next investigated the sensitivity of SAMHD1 structure, function, and regulation to replacements at positions under positive selection surrounding the phosphorylation site. We identified 4 sites of particular interest due to their structural context, proximity to T592, and the distribution of residues seen in different species (Fig. 2, yellow). Site 566 (arginine in humans) has undergone many physicochemically nonconservative substitutions to residues greatly varying in size, such as glycine, glutamine, and histidine, despite being mostly buried (*SI Appendix, Fig. S5*). Site 574 (alanine in humans) is almost entirely buried in the C-terminal region, yet residues with much larger side chains and contrasting physicochemical properties were observed in other species, including leucine, phenylalanine, and serine (*SI Appendix, Fig. S6*). Sites 594 and 596 (glutamine and lysine in humans, respectively) are both very close to T592 (Fig. 2B), and, again, chemically diverse residues were observed at these positions (*SI Appendix, Figs. S7 and S8*). The amino acid at site 596 marks the divergence of New World monkeys (negatively charged aspartate) from Old World monkeys and apes (positively charged lysine). To investigate whether these residues are likely to have the same structural context across species, we used homology modeling to predict animal SAMHD1 protein structures (38). Comparing each animal model structure to the human structure, the minimum root mean square deviation (rmsd) of atoms was below 0.4 Å, and corresponding Z-scores with respect to random structure alignments were all above 50, indicating very close structural similarity (*SI Appendix, Table S6*; ref. 39).

We then selected 1 or more residues that had been observed in multiple animal species at these 4 sites and introduced them into a human SAMHD1 background to test effects on function. Further homology-modeling experiments had shown that these residues had almost identical position and orientation when comparing mutant human SAMHD1 model structures and animal model structures for which these residues are wild type (WT), as measured by rmsd between atoms in their side chains; this suggested that the mutant residues in the human background accurately recapitulate their situation in their respective WT animal SAMHD1 structures (*SI Appendix, Table S7 and Fig. S9*). WT and mutant SAMHD1 constructs were expressed in *Escherichia coli*, and the activator and substrate dependence of tetramerization of purified proteins was analyzed by size-exclusion chromatography coupled to multiangle laser light scattering (SEC-MALLS). Each mutant was tested for tetramerization of the apo protein and with the addition of either GTP activator alone or GTP and additional dATP substrate. As with WT human SAMHD1, these experiments showed that no human SAMHD1 mutants tetramerize in the apo form or with the addition of GTP alone (*SI Appendix, Fig. S10*). On incubation with GTP and substrate dATP, all of the mutants underwent tetramerization with similar efficiency to WT, suggesting that they were not positively selected for this purpose (*SI Appendix, Fig. S11*). The degrees of phosphorylation of these variants were examined by Western blot using phospho-specific SAMHD1 antibodies (*SI Appendix, Fig. S12*). Variants Q594L and Q594R showed reduced signal relative to total SAMHD1 levels, suggesting that at least some positively selected sites do impinge on phosphorylation; however, differential binding affinity of the antibody for the mutated sites cannot be ruled out. Overall, introduction of these mutations did not seem to alter the protein expression levels greatly (*SI Appendix, Fig. S12*).

We next assessed the triphosphohydrolase activities of the purified mutant enzymes. All SAMHD1 mutants had reduced (up

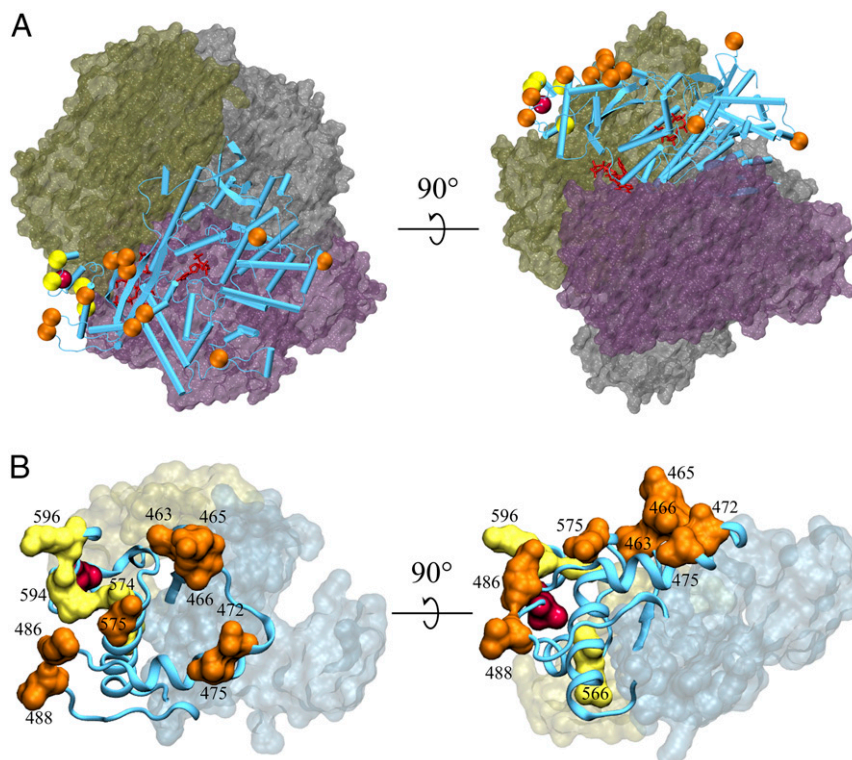


Fig. 2. Sites identified as under positive selection in mammalian SAMHD1, shown on a published crystal structure (Protein Data Bank ID code 4TNP; ref. 25; residues 114–276 and 282–599). (A) Human SAMHD1 tetramer with foreground monomer shown as cyan cylindrical-helix cartoon and background monomers shown as colored surface representations. The 15 sites both under positive selection and present in this structure have their α -carbons shown as orange or yellow spheres, where yellow indicates sites in the C-terminal region chosen for point mutation experiments; threonine 592 is similarly shown and colored red. dNTP molecules bound in the foreground monomer are shown as red sticks. (B) Enlarged view of the C-terminal region, with foreground monomer shown as a plain cartoon, all atoms of sites under positive selection shown by surface representation, and background monomers as translucent surfaces. Highlighted sites are annotated with human sequence numbering, and site colors are as in A. (B, Left) Site 566 is obscured by site 575. (B, Right) Sites 574 and 594 are obscured by sites 575 and 486, respectively.

to 2.5-fold) steady-state catalytic rates relative to WT, potentially due to a mismatch between the new residue and the human background sequence, but enzymatic activity was not severely disrupted (Fig. 3 A and B). The greatest effect was with mutant K596D, which, as mentioned above, marks the distinction between Old and New World primates (*SI Appendix*, Fig. S8). All mutants had similar K_M values to the WT, with mutations causing both increases and decreases in this parameter. One mutant, K596M, had a 3-fold reduction in K_M , suggesting significantly enhanced substrate binding.

We then assessed the effect of changes at these sites under positive selection on the ability of SAMHD1 to restrict HIV-1 infection, measured in differentiated U937 cells using our previously described 2-color flow-cytometry HIV-1 restriction assay (23, 40). HIV-1 restriction was not ablated in any of the mutants tested, but appeared less efficient for several mutants, consistent with their observed reduction in enzymatic activity (Fig. 3C). Of note, a few variants (R566Q, Q594L, and Q596P) restricted HIV-1 infection as efficiently as WT, despite some reduction of enzymatic activity. We expanded our analysis to other lentiviruses, equine infectious anemia virus (EIAV) and FIV, comparing the effects of substitutions A574L (found in both horse and cat), Q594R (cat), Q594L (horse) (*SI Appendix*, Figs. S6 and S7), and WT human SAMHD1 on virus restriction (Fig. 3 D and E). Notably, variants found in cat conferred stronger restriction against FIV and EIAV than HIV-1, whereas the horse variant was relatively weaker, which paralleled their enzymatic activity, with Q594L being rather less active than A574L or Q594R (Fig. 3 A and B). Interestingly, the reverse was

also true, as Q594L conferred strong restriction of HIV-1, but weaker EIAV/FIV restriction. Taken together, these results demonstrate that positively selected sites around T592 modulate enzyme activity, lentiviral restriction, and phosphorylation and provide insights into SAMHD1 regulation and evolution.

Discussion

Viruses impose a significant burden upon their hosts, forcing them into persistent evolutionary conflict. Host antiviral responses must be carefully modulated, however, to prevent self-damage or dysregulation of pathways critical for cell homeostasis. There is evidence that such modulation can be achieved by controlling levels of transcription of innate immunity genes (41), but it has been unclear whether regulation of the activity of innate immunity proteins is also important. By investigating the deep evolutionary history of the innate immune protein SAMHD1, our results suggest that this mechanism might indeed be important and afford appropriate modulation.

To date, SAMHD1 evolution has been closely studied in primate lentiviruses and their hosts, where the characteristic signatures of positive diversifying selection have been observed, specifically at the N- and C-terminal regions that interact with Vpr/Vpx viral proteins (17, 19). It has remained unclear whether this coevolution has been restricted to primates and whether the activity of SAMHD1 might be modulated by the host in response to pathogens. We have found that positive selection in SAMHD1 is not restricted to primates, but rather has been pervasive throughout mammalian evolution and can be observed in every mammal group for which data are available (Fig. 1). This reflects

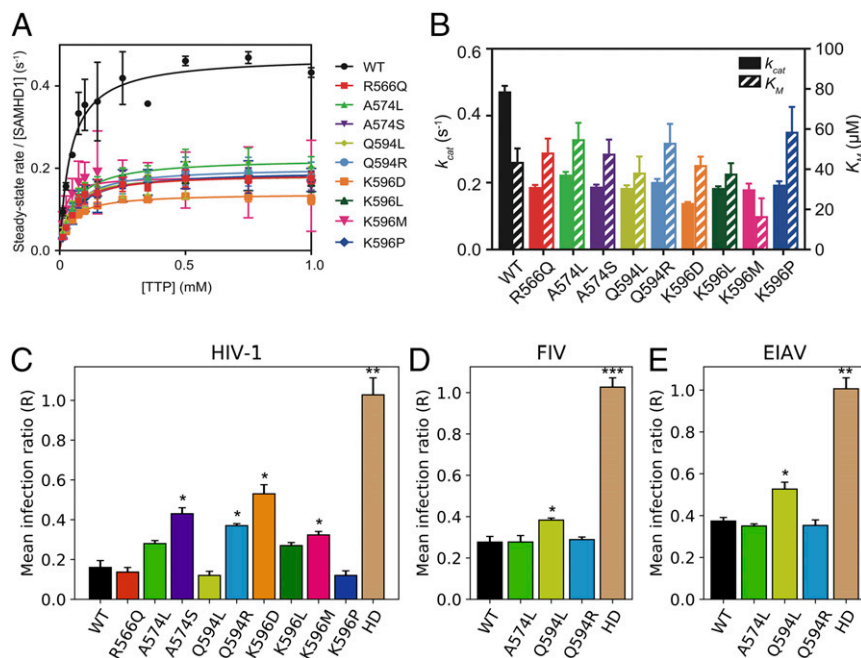


Fig. 3. Activities of SAMHD1 positive selection point mutants. (A) Steady-state kinetics of triphosphohydrolase activity of human SAMHD1 WT and point mutants. Plots show the concentration dependence of the rate of SAMHD1 TTP hydrolysis in the presence of 0.2 mM GTP activator. The error bars are the SEM from triplicate measurements. (B) Kinetic parameters k_{cat} and K_M obtained from Michaelis–Menten analysis of data in A. Error bars are SEM from triplicate measurements. (C) Anti-HIV-1 restriction activity of human SAMHD1 WT and point mutants. The bars show the infection ratio of differentiated U937 cells expressing each variant with respect to an untransduced control. HD is the negative control human SAMHD1 active-site knockout mutant HD206/207AA. Error bars are the SEM from $n = 3$ or $n = 4$ measurements; asterisks indicate statistically significant difference from WT measured by Student's t test and controlling false discovery rate by the Benjamini–Hochberg method. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. (D) Anti-FIV restriction activity, as in C. (E) Anti-EIAV restriction activity, as in C.

either widespread similar adaptation to recent pathogens in a wide range of mammals or, perhaps more likely, an ancient and ongoing battle between the mammalian innate immune system and viruses, potentially involving other restriction factors and taxonomic groups.

We found that regions of SAMHD1 are under positive selection in multiple mammal groups, such as the C-terminal Vpx/Vpr binding site, which is under positive selection in Glires, Cetartiodactyla, Chiroptera, and Other Mammals (Fig. 1 and *SI Appendix, Fig. S4*). The surfaces of primate SAMHD1 bound by Vpx/Vpr appear to vary and even fluctuate throughout evolutionary time (18), but the density of sites identified in this region suggests that viral antagonists in other species may also target this region. The signatures of positive selection in mammalian SAMHD1 may represent both ongoing evolutionary conflict and a record of past battles with viruses now extinct.

Several classes of DNA-producing viruses are restricted in the presence of SAMHD1 (3–10), suggesting that many diverse animal viruses have an evolutionary incentive to overcome its effects. SAMHD1 sensitivity extends to FIV and EIAV, although there is no evidence of SAMHD1 antagonism by these or other lentiviruses, suggesting that, like HIV-1, their replication strategies do not require it (10). Nonetheless, ancient endogenous retrovirus sequences have been found in diverse species genomes (42), indicating that retroviruses have coevolved with animals and their restriction factors throughout metazoan evolution and may have imposed selective pressure on their hosts' SAMHD1.

We identified a significant clustering of sites under positive selection in the domain containing the phosphorylation site T592, in mammals in general as well as in Cetartiodactyla, Chiroptera, and Other Mammals. Replacing human SAMHD1 residues at sites under positive selection in this domain with those found in other species—thereby mimicking possible residue substitutions

due to positive selection—modestly enhanced tetramerization and substrate binding (K_M), but reduced the catalytic rate (k_{cat}) (*SI Appendix, Fig. S11* and Fig. 3 A and B). In the case of the Q594L and Q594R variants, SAMHD1 phosphorylation also appeared to be reduced (*SI Appendix, Fig. S12*). Several of the mutations that decreased enzymatic activity also reduced restriction of HIV-1 (Fig. 3 A–C), demonstrating that sites distant from the active site and under positive selection have important influence on SAMHD1 function. Furthermore, we observed that substitutions normally found in cat and horse resulted in loss of anti-HIV-1 activity, but maintained restriction of FIV and EIAV, suggesting changes in SAMHD1 antiviral specificity over evolutionary time. These results suggest that positive selection in the regions we observed involves the modulation of SAMHD1's activity, which might be differentially adapted in different species, perhaps due to the different biological properties of reverse transcriptase (RT) from these lentiviruses (43, 44). More generally, adaptation to each host's range of pathogens may therefore have driven the widespread positive selection we observe.

There is an interesting parallel with the IFN-induced transmembrane 3 protein (IFITM3), which restricts diverse enveloped viruses. Several posttranslational modifications are determined by IFITM3's N terminus, and replacing residues here with those found in nonhuman primate orthologs resulted in opposite effects on restriction of different viruses, similarly suggesting a trade-off in viral specificity (45). Meanwhile, SAMHD1 is in contrast with the rodent transferrin receptor (TfR1), which has undergone positive selection in regions bound by viruses to mediate cell entry: Mutations at these positions disrupt TfR1-dependent virus entry, but without influencing TfR1's core functions involved with iron transport, suggesting effective separation between selection pressures to avoid pathogen interaction and retain biological activity (46).

One possible explanation is positive diversifying selection on SAMHD1's function as a dNTP regulator, independent of its role as a restriction factor. However, it is not clear why an enzyme fulfilling an important homeostatic function would be under diversifying selection, and, moreover, there is no known precedent for diversifying selection to be associated with enzyme activity/regulation. Alternatively, substitutions driven by selection pressure from viruses might impact SAMHD1 enzymatic activity, resulting in selection for compensatory substitutions. The surface surrounding T592 may itself constitute an interaction interface, meaning that positive selection in this domain has occurred to disrupt interaction with viral proteins. However, the sites identified as under positive selection do not form an obvious binding surface, since some are buried. A third possibility is that SAMHD1 regulation is directly involved in the virus–host evolutionary conflict, adapting catalytic efficiency or propensity for (de)phosphorylation to regulate antiviral potency. For instance, sites 618 and 619, which were identified as under positive selection in mammals, nonprimates, and individually in Carnivora (site 618) and Chiroptera (site 619), have been implicated in binding with cyclin A2, an interaction responsible for the phosphorylation of SAMHD1 T592 (28).

In this study, we have found evidence of an arms race between viruses and their hosts of a significantly larger scale than previously realized. Far from being limited to primates, we have found that positive selection has occurred throughout the evolution of SAMHD1 in mammals, most likely due to selection pressure applied by modern or extinct DNA-producing viruses. Significantly, we have found evidence that this adaptation can involve modulation of a host protein's function. This suggests a model in which SAMHD1 is under selective constraints to both conserve its metabolic and antiviral functions while continually adapting to pressures imposed by viruses. This principle of balancing function conservation with continual adaptation is likely to apply to other innate immunity components, as they adapt to manage the cost of evolutionary conflict.

Methods

Licensing for Human or Animal Materials. This work involved no experiments using human or animal materials.

Sequence Data and Alignment. Mammalian SAMHD1 DNA sequences were collected by using NCBI BLAST (blastn algorithm) with human SAMHD1 coding sequence (accession no. NM_015474.3) as the query. Mostly, these were predicted mRNA sequences originating from automated analysis of mammal genome sequences, while the majority of primate sequences originated from previous SAMHD1 studies (17, 19). Where more than 1 sequence was available from a single species (usually predicted transcript variants), sequences most closely matching the human sequence were selected. The sequence for Tasmanian devil (*Sarcophilus harrisi*) was found to be divided into 2 sequence records (accession nos. XM_003758997.2 and XM_012553363.1); these were concatenated to give a full-length sequence. Preliminary phylogenetic analysis including nonmammal taxa found the platypus (*Ornithorhynchus anatinus*) sequence incorrectly positioned outside of the mammalian clade; this was therefore excluded from subsequent analysis. The list of species and accession numbers for sequences used are listed in *SI Appendix, Table S1*. Sequences that were <70% of the length of the human SAMHD1 sequence were excluded.

The nucleotide coding sequences were initially aligned as translated protein by using MUSCLE [Version 3.8.31; ref. 47, as implemented in SEAVIEW (Version 4.4.0); ref. 48] and then further edited manually, with a highly conservative approach: Sections within sequences which could not be aligned with high confidence were masked, such that they would be treated as missing data (equivalent to alignment gaps) by phylogenetics tools. Alignment columns containing no data (exclusively gaps or masked codons) in 20% of sequences were removed.

Phylogeny Estimation. A phylogenetic tree was estimated by maximum likelihood using RAxML (Version 7.7.2; refs. 49 and 50) with the general time-reversible substitution model and gamma-distributed rate heterogeneity. Confidence in the tree topology was assessed by estimating trees from 1,000

nonparametric bootstrap samples (51). In repeating selection analyses with alternative tree topologies, we used 3 of the nonparametric bootstrap trees generated by RAxML that had independent maximum parsimony trees as initial estimates. Tree figures were produced by using FigTree (Version 1.3.1) (52).

Selection Analysis. Selection analyses were performed by using the codeml program of the PAML package (Version 4.7a; ref. 34). We used the site-specific tests of positive selection M1a/M2a (32) and M7/M8 (33). For the large all-mammals dataset, the more complex M8 model was unable to converge, and for mammal subgroups, sites identified with M2a were almost always subsets of those identified with M8.

To reduce the risk of the log-likelihood optimization reaching a local optimum, all program runs were performed 5 times with different initial parameters for the transition/transversion ratio (κ) and dN/dS ratio (ω): 0.1, 1, and 10. Tree branch lengths were first optimized with codeml's model 0 (which allows a single ω value) with the corresponding initial parameter values, and these branch lengths were used as starting values in subsequent analyses with more complex models. Codon stationary frequencies were estimated by using the F1x4 model by default, but F3x4 was used if optimization difficulties were encountered with F1x4 (null and alternative models were always compared by using consistent codon frequency models; *SI Appendix, Table S2*).

Statistical justification for the alternative model was assessed by using the likelihood-ratio test, using 2 degrees of freedom. Sites were identified as being under positive selection if the computed Bayes empirical Bayes probability for the site belonging to the positive selection class was >0.95.

Clustering of Sites under Positive Selection. Statistical significance for sites under positive selection clustering in regions of the SAMHD1 linear sequence was determined by combinatorial analysis. Briefly, in a sequence of $N = 626$ sites, of which n are under positive selection, there is a region of biological interest comprising R sites ($R < N$), of which r are under positive selection ($r \leq n$). The probability of r or more of the n positively selected sites occurring in this region at random is given by

$$\frac{\sum_{s \geq r} \binom{R}{s} \binom{N-R}{n-s}}{\binom{N}{n}}$$

Protein Expression and Purification. The DNA sequence for human SAMHD1 (residues M1–M626) was amplified by PCR from plasmid template and inserted into a pET52b expression vector (Novagen) using ligation independent cloning to produce an N-terminal StrepII-tag fusion protein. Point mutations corresponding to the residues found in other species were introduced into the WT protein construct by using the Quikchange II kit. All insert sequences were verified by DNA sequencing. Strep-tagged SAMHD1 constructs were expressed in the *E. coli* strain Rosetta 2 (DE3) grown at 37 °C with shaking. Protein expression was induced by addition of 0.1 mM isopropyl β -D-thiogalactopyranoside (IPTG) to log-phase cultures ($A_{600} = 0.5$), and the cells were incubated for a further 20 h at 18 °C. Cells were harvested by centrifugation resuspended in 30 mL of lysis buffer [50 mM Tris-HCl pH 7.8, 500 mM NaCl, 4 mM $MgCl_2$, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP), 1 \times EDTA-free mini complete protease inhibitors (Roche), and 0.1 U/mL Benzozase (Novagen)] per pellet of 1 L of bacterial culture and lysed by disruption in EmulsiFlex-C5 homogenizer (Avestin). The lysate was cleared by centrifugation for 1 h at 48,000 $\times g$ and 4 °C, then applied to a 10-mL StrepTactin affinity column (IBA) followed by 600 mL of wash buffer (50 mM Tris-HCl, pH 7.8, 500 mM NaCl, 4 mM $MgCl_2$, and 0.5 mM TCEP) at 4 °C. Bound proteins were eluted from the column by circulation of 1 mg of 3C protease (GE) in 10 mL of wash buffer over the column in a closed circuit overnight. The 3C protease was removed by incubation of the eluent with 500 μ L of GSH-Sepharose (GE). After centrifugation to remove the resin, the supernatant was concentrated to 5 mL and applied to a Superdex 200 16/60 (GE) size-exclusion column equilibrated with 10 mM Tris-HCl (pH 7.8), 150 mM NaCl, 4 mM $MgCl_2$, and 0.5 mM TCEP. Peak fractions were concentrated to ~20 mg/mL and flash-frozen in liquid nitrogen in small aliquots.

SEC-MALLS. SEC-MALLS was used to determine the molar mass composition of SAMHD1 samples upon addition of deoxynucleotide/nucleotide substrates (dATP; 500 μ M) and activators (GTP; 200 μ M). Samples (30 μ M SAMHD1 and variants) were incubated with substrate and activator at 4 °C for 5 min, and then 100 μ L was applied to a Superdex 200 10/300 GL column equilibrated in 20 mM Tris-HCl, 150 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, and 3 mM $NaNO_3$ (pH 8.0) at a flow rate of 0.5 mL/min. The scattered light intensity and

protein concentration of the column eluate were recorded by using a DAWN-HELEOS laser photometer and an OPTILAB-REX differential refractometer (dRI) ($dn/dc = 0.186$), respectively. The weight-averaged molecular mass of material contained in chromatographic peaks and peak integrals were determined by using the combined data from both detectors in the ASTRA software (Version 6.1; Wyatt Technology Corp.).

SAMHD1 Catalytic Activity Assays. The hydrolysis of dNTPs by SAMHD1 and SAMHD1 variants was quantified by using a coupled assay utilizing the biosensor MDCC-PBP (53, 54) to measure phosphate release from combined SAMHD1 triphosphohydrolase and *Saccharomyces cerevisiae* Ppx1 exopolyphosphatase activity as described (55). In a typical experiment, solutions containing SAMHD1 constructs, Ppx1, MDCC-PBP, and GTP were incubated for 5 min in assay buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 5 mM MgCl₂, and 2 mM TCEP) at 25 °C before the reaction was initiated by the addition of substrate [thymidine triphosphate (TTP)]. The final concentrations were 100 nM SAMHD1, 10 nM Ppx1, 40 μM MDCC-PBP, 0.2 mM GTP, and varying concentration of TTP. The fluorescence intensity was recorded at 430-nm excitation and 465-nm emission over a period of 10–30 min in a Clariostar multiwell plate reader (BMG). Steady-state rates were obtained from time courses of P_i formation by linear regression of the data points in the linear phase of the reaction. Rates were divided by the SAMHD1 concentration and plotted versus substrate concentration. Apparent dissociation constants for substrate binding (K_M) and catalytic constants (k_{cat}) were then determined by nonlinear least-squares fitting using either a hyperbolic or a Hill function. All measurements were performed in duplicate or triplicate.

SAMHD1 Plasmids. WT SAMHD1 was previously cloned into the pLGateway-YFP vector to create pLGateway_SAMHD1IRESYFP (23). SAMHD1 mutants (R566Q, A574S, A574L, Q594L, Q594R, K596D, K596P, K596L, and K596M) were created by PCR-based (Pfu; Agilent) site-directed mutagenesis of the pLGateway_SAMHD1IRESYFP vector using the manufacturer's protocol and the primers listed in *SI Appendix, Table S8*.

Cells and Virus Production. The 293T cells were maintained in Dulbecco's modified Eagle medium (Life Technologies) and U937 cells in RPMI +[L]-glutamine (Life Technologies), each supplemented with 10% heat-inactivated fetal bovine serum (Labtech) and 1% penicillin/streptomycin (Sigma). Moloney murine leukemia virus-based SAMHD1 and yellow fluorescent protein (YFP)-expressing virus-like particles (VLPs) were produced by cotransfecting 293T cells with pVSV-G (56), pKB4 (57), and pLGateway_SAMHD1IRESYFP (WT or mutants) into 293T cells and harvesting 48 h after transfection. HIV-1-GFP was made by cotransfection of pVSV-G, p8.91 (58), and pCSGWGFP (59). EIAV-GFP vector was made with plasmids pEIAV-SIN6.1 CGFPW, pEV53D (gift of John Olsen, Addgene, Chapel Hill, NC; ref. 60), and pVSV-G. FIV-GFP vector was made by using plasmids pGIN SIN FIVGFP, pFP93 (gift of Eric Poeschla, University of Colorado, Denver), and pVSV-G (61). Plasmids were mixed at ratio 3:3:3 μg/10-cm dish. Supernatants were collected 48 and 72 h post-transfection, centrifuged at 3,000 rpm, filtered through a 0.45-μm filter, and

concentrated on the same day by centrifugation at 100,000 × g for 2 h at 4 °C through a 20% sucrose cushion before resuspension in serum-free RPMI. VLPs were titered on U937 cells for normalization prior to infection. Viral stocks were stored at –80 °C until use.

HIV-1 Restriction Assay. Undifferentiated U937 cells (3×10^5 cells in 12-well plate) were transduced with SAMHD1-YFP VLPs by spinoculation at 800 × g for 90 min in the presence of 10 μg/mL polybrene. After 72 h incubation, the cells were passaged 1:4 and differentiated with 100 nM phorbol 12-myristate 13-acetate (Sigma) for 96 h. Differentiated cells were infected in triplicate with HIV-1-GFP in the presence of 10 μg/mL polybrene, and restriction was assessed after 72 h by 2-color flow cytometry using a Fortessa x20 analyzer (BD Biosciences). Data were analyzed by using the FlowJo software suite. Restriction was calculated by dividing the percentage of SAMHD1-expressing (YFP +ve) cells that were infected with HIV-1 (GFP +ve) by the percentage of SAMHD1-negative cells that were infected to give an infection ratio, *R*.

Immunoblotting. As described above, SAMHD1-YFP-expressing U937 cells were sorted on MoFlo XDP (Beckman Coulter). After culturing for 3–4 d to recover from cell sorting, 1×10^7 cells were harvested and washed twice with phosphate-buffered saline. Supernatant was removed, and cells were lysed with 100 μg of radioimmunoprecipitation assay buffer (ThermoFisher) in the presence of protease inhibitors (Roche), DNase (Invitrogen), and phosphatase inhibitor (ThermoFisher) at 4 °C for 1 h. The lysates were subsequently sonicated for 5 min, ×2 pulses, and 40% amplitude and centrifuged for 30 min at 4 °C and 48,000 × g. Laemmli buffer (2×, Sigma) was added at a 1:1 ratio to the lysate and boiled at 95 °C for 5 min. To separate proteins, samples were loaded on sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred onto polyvinylidene fluoride membranes (Millipore), which were then incubated with blocking buffer (5% bovine serum albumin and 0.05% Tween 20 in Tris-HCl buffer, pH 7.5; Sigma) at 4 °C overnight. Mouse anti-SAMHD1 1F6 monoclonal antibody (GeneTex; 1:500) and rabbit anti-phosphorylated-SAMHD1 monoclonal antibody (CST; 1:1,000) were used as primary antibodies. Corresponding secondary antibodies used were goat anti-mouse IRDye800CW (LI-COR; 1:10,000) and goat anti-rabbit IRDye800CW (LI-COR; 1:10,000). Fluorescence was measured by using the LI-COR Odyssey imaging (LI-COR Bioscience).

ACKNOWLEDGMENTS. This work was supported by UK Medical Research Council Grants MC_U117573805 (to C.M., G.T., and R.A.G.) and G0900950 (to A.F.); UK Biotechnology and Biological Sciences Research Council Grant BB/P007562/1 (to R.A.G.); European Union Framework Program 7 HivinNov Grant 305137 (to A.F.); Wellcome Trust Grants 108014Z/15Z (to E.R.M. and I.A.T.) and 108012Z/15Z (to B.S. and J.P.S.); and the Francis Crick Institute (which receives its core funding from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust) Grants FC001042 (to K.N.B.), FC001162 (to J.P.S.), and FC001178 (to I.A.T.). C.R. is supported by a studentship from University College London CoMPLEX. We thank David D. Pollock and 2 anonymous reviewers for comments on the manuscript.

1. F. Kirchhoff, Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. *Cell Host Microbe* **8**, 55–67 (2010).
2. M. H. Malim, M. Emerman, HIV-1 accessory proteins—Ensuring viral survival in a hostile environment. *Cell Host Microbe* **3**, 388–398 (2008).
3. K. Hrecka *et al.*, Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* **474**, 658–661 (2011).
4. N. Laguette *et al.*, SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* **474**, 654–657 (2011).
5. T. Gramberg *et al.*, Restriction of diverse retroviruses by SAMHD1. *Retrovirology* **10**, 26 (2013).
6. J. Choi, J. Ryoo, C. Oh, S. Hwang, K. Ahn, SAMHD1 specifically restricts retroviruses through its RNase activity. *Retrovirology* **12**, 46 (2015).
7. J. A. Hollenbaugh *et al.*, Host factor SAMHD1 restricts DNA viruses in non-dividing myeloid cells. *PLoS Pathog.* **9**, e1003481 (2013).
8. E. T. Kim, T. E. White, A. Brandariz-Núñez, F. Diaz-Griffero, M. D. Weitzman, SAMHD1 restricts herpes simplex virus 1 in macrophages by limiting DNA replication. *J. Virol.* **87**, 12949–12956 (2013).
9. A. F. R. Sommer *et al.*, Restrictive influence of SAMHD1 on hepatitis B virus life cycle. *Sci. Rep.* **6**, 26616 (2016).
10. S. A. Meredy *et al.*, Interplay of ancestral non-primate lentiviruses with the virus-restricting SAMHD1 proteins of their hosts. *J. Biol. Chem.* **293**, 16402–16412 (2018).
11. D. C. Goldstone *et al.*, HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature* **480**, 379–382 (2011).
12. H. Lahouassa *et al.*, SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.* **13**, 223–228 (2012).
13. S. Srivastava *et al.*, Lentiviral Vpx accessory factor targets VprBP/DCAF1 substrate adaptor for cullin 4 E3 ubiquitin ligase to enable macrophage infection. *PLoS Pathog.* **4**, e1000059 (2008).
14. A. Bergamaschi *et al.*, The human immunodeficiency virus type 2 Vpx protein usurps the CUL4A-DDB1 DCAF1 ubiquitin ligase to overcome a postentry block in macrophage infection. *J. Virol.* **83**, 4854–4860 (2009).
15. C. Goujon *et al.*, SIVSM/HIV-2 Vpx proteins promote retroviral escape from a proteasome-dependent restriction pathway present in human dendritic cells. *Retrovirology* **4**, 2 (2007).
16. C. Goujon *et al.*, Characterization of simian immunodeficiency virus SIVSM/human immunodeficiency virus type 2 Vpx function in human myeloid cells. *J. Virol.* **82**, 12335–12345 (2008).
17. E. S. Lim *et al.*, The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host Microbe* **11**, 194–204 (2012).
18. O. I. Fregoso *et al.*, Evolutionary toggling of Vpx/Vpr specificity results in divergent recognition of the restriction factor SAMHD1. *PLoS Pathog.* **9**, e1003496 (2013).
19. N. Laguette *et al.*, Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. *Cell Host Microbe* **11**, 205–217 (2012).
20. C. J. Spragg, M. Emerman, Antagonism of SAMHD1 is actively maintained in natural infections of simian immunodeficiency virus. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 21136–21141 (2013).
21. E. Franzolin *et al.*, The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14272–14277 (2013).

22. G. I. Rice *et al.*, Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nat. Genet.* **41**, 829–832 (2009).
23. L. H. Arnold *et al.*, Phospho-dependent regulation of SAMHD1 oligomerisation couples catalysis and restriction. *PLoS Pathog.* **11**, e1005194 (2015).
24. J. Yan *et al.*, Tetramerization of SAMHD1 is required for biological activity and inhibition of HIV infection. *J. Biol. Chem.* **288**, 10406–10417 (2013).
25. X. Ji, C. Tang, Q. Zhao, W. Wang, Y. Xiong, Structural basis of cellular dNTP regulation by SAMHD1. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4305–E4314 (2014).
26. A. Cribier, B. Descours, A. L. Valadão, N. Laguette, M. Benkirane, Phosphorylation of SAMHD1 by cyclin A2/CDK1 regulates its restriction activity toward HIV-1. *Cell Rep.* **3**, 1036–1043 (2013).
27. E. Pauls *et al.*, Cell cycle control and HIV-1 susceptibility are linked by CDK6-dependent CDK2 phosphorylation of SAMHD1 in myeloid and lymphoid cells. *J. Immunol.* **193**, 1988–1997 (2014).
28. J. Yan *et al.*, CyclinA2-Cyclin-dependent kinase regulates SAMHD1 protein phosphorylation domain. *J. Biol. Chem.* **290**, 13279–13292 (2015).
29. S. Wittmann *et al.*, Phosphorylation of murine SAMHD1 regulates its antiretroviral activity. *Retrovirology* **12**, 103 (2015).
30. F. Wang *et al.*, Phosphorylation of mouse SAMHD1 regulates its restriction of human immunodeficiency virus type 1 infection, but not murine leukemia virus infection. *Virology* **487**, 273–284 (2016).
31. R. W. Meredith *et al.*, Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
32. R. Nielsen, Z. Yang, Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929–936 (1998).
33. Z. Yang, R. Nielsen, N. Goldman, A. M. Pedersen, Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449 (2000).
34. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
35. R. M. McBee, S. A. Rozmiarek, N. R. Meyerson, P. A. Rowley, S. L. Sawyer, The effect of species representation on the detection of positive selection in primate gene data sets. *Mol. Biol. Evol.* **32**, 1091–1096 (2015).
36. D. Schwefel *et al.*, Structural basis of lentiviral subversion of a cellular protein degradation pathway. *Nature* **505**, 234–238 (2014).
37. D. Schwefel *et al.*, Molecular determinants for recognition of divergent SAMHD1 proteins by the lentiviral accessory protein Vpx. *Cell Host Microbe* **17**, 489–499 (2015).
38. M. Biasini *et al.*, SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252–8 (2014).
39. L. Holm, L. M. Laakso, Dali server update. *Nucleic Acids Res.* **44**, W351–5 (2016).
40. P. Ordonez *et al.*, SAMHD1 enhances nucleoside-analogue efficacy against HIV-1 in myeloid cells. *Sci. Rep.* **7**, 42824 (2017).
41. T. Hagai *et al.*, Gene expression variability across cells and species shapes innate immunity. *Nature* **563**, 197–202 (2018).
42. R. Gifford, M. Tristem, The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**, 291–315 (2003).
43. T. Rubinek *et al.*, The catalytic properties of the reverse transcriptase of the lentivirus equine infectious anemia virus. *Eur. J. Biochem.* **219**, 977–983 (1994).
44. M. Amacker, U. Hübscher, Chimeric HIV-1 and feline immunodeficiency virus reverse transcriptases: Critical role of the p51 subunit in the structural integrity of heterodimeric lentiviral DNA polymerases. *J. Mol. Biol.* **278**, 757–765 (1998).
45. A. A. Compton *et al.*, Natural mutations in IFITM3 modulate post-translational regulation and toggle antiviral specificity. *EMBO Rep.* **17**, 1657–1671 (2016).
46. A. Demogines, J. Abraham, H. Choe, M. Farzan, S. L. Sawyer, Dual host-virus arms races shape an essential housekeeping protein. *PLoS Biol.* **11**, e1001571 (2013).
47. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
48. M. Gouy, S. Guindon, O. Gascuel, SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
49. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).
50. M. Ott, J. Zola, A. Stamatakis, S. Aluru, “Large-scale maximum likelihood-based phylogenetic analysis on the IBM BlueGene/L” in *Proceedings of the 2007 ACM/IEEE Conference on Supercomputing (SC '07)* (Association for Computing Machinery, New York, 2007), pp. 1–11.
51. A. Stamatakis, P. Hoover, J. Rougemont, A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.* **57**, 758–771 (2008).
52. A. Rambaut, FigTree, Version 1.4.4. <https://github.com/rambaut/figtree/releases>. Accessed 1 December 2018.
53. M. Brune, J. L. Hunter, J. E. Corrie, M. R. Webb, Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase. *Biochemistry* **33**, 8262–8271 (1994).
54. M. Brune *et al.*, Mechanism of inorganic phosphate interaction with phosphate binding protein from *Escherichia coli*. *Biochemistry* **37**, 10370–10380 (1998).
55. L. H. Arnold, S. Kunzelmann, M. R. Webb, I. A. Taylor, A continuous enzyme-coupled assay for triphosphohydrolase activity of HIV-1 restriction factor SAMHD1. *Antimicrob. Agents Chemother.* **59**, 186–192 (2015).
56. M. Bock, K. N. Bishop, G. Towers, J. P. Stoye, Use of a transient assay for studying the genetic determinants of Fv1 restriction. *J. Virol.* **74**, 7422–7430 (2000).
57. H. C. T. Groom *et al.*, Absence of xenotropic murine leukaemia virus-related virus in UK patients with chronic fatigue syndrome. *Retrovirology* **7**, 10 (2010).
58. L. Naldini *et al.*, In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science* **272**, 263–267 (1996).
59. J. W. Bainbridge *et al.*, In vivo gene transfer to the mouse eye using an HIV-based lentiviral vector; efficient long-term transduction of corneal endothelium and retinal pigment epithelium. *Gene Ther.* **8**, 1665–1668 (2001).
60. J. C. Olsen, Gene transfer vectors derived from equine infectious anemia virus. *Gene Ther.* **5**, 1481–1487 (1998).
61. D. T. Saenz, R. Barraza, N. Loewen, W. Teo, E. M. Poeschla, Feline immunodeficiency virus-based lentiviral vectors. *Cold Spring Harb. Protoc.* **2012**, 71–76 (2012).