

DNA Editing of LTR Retrotransposons Reveals the Impact of APOBECs on Vertebrate Genomes

Binyamin A. Knisbacher¹ and Erez Y. Levanon^{*,1}

¹The Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan, 52900, Israel

*Corresponding author: E-mail: erez.levanon@biu.ac.il.

Associate editor: Sudhir Kumar

Abstract

Long terminal repeat retrotransposons (LTR) are widespread in vertebrates and their dynamism facilitates genome evolution. However, these endogenous retroviruses (ERVs) must be restricted to maintain genomic stability. The APOBECs, a protein family that can edit C-to-U in DNA, do so by interfering with reverse transcription and hypermutating retrotransposon DNA. In some cases, a retrotransposon may integrate into the genome despite being hypermutated. Such an event introduces a unique sequence into the genome, increasing retrotransposon diversity and the probability of developing new function at the locus of insertion. The prevalence of this phenomenon and its effects on vertebrate genomes are still unclear. In this study, we screened ERV sequences in the genomes of 123 diverse species and identified hundreds of thousands of edited sites in multiple vertebrate lineages, including placental mammals, marsupials, and birds. Numerous edited ERVs carry high mutation loads, some with greater than 350 edited sites, profoundly damaging their open-reading frames. For many of the species studied, this is the first evidence that APOBECs are active players in their innate immune system. Unexpectedly, some birds and especially zebra finch and medium ground-finch (one of Darwin's finches) are exceptionally enriched in DNA editing. We demonstrate that edited retrotransposons may be preferentially retained in active genomic regions, as reflected from their enrichment in genes, exons, promoters, and transcription start sites, thereby raising the probability of their exaptation for novel function. In conclusion, DNA editing of retrotransposons by APOBECs has a substantial role in vertebrate innate immunity and may boost genome evolution.

Key words: DNA editing, retrotransposons, *apobec3*, *apobec*, genome evolution, innate immunity.

Introduction

The AID/APOBECs are a vertebrate-specific family of cytidine deaminases involved in numerous biological pathways (reviewed in Conticello et al. 2007; Holmes et al. 2007; Koito and Ikeda 2011, 2013; Smith et al. 2012; Refsland and Harris 2013; Prohaska et al. 2014). Through their ability to convert cytosines to uracils (C-to-U) in DNA, a process coined "DNA editing," the family plays a key role in both arms of the immune system. AID induces antibody diversification and maturation by somatic hypermutation in adaptive immunity (Muramatsu et al. 2000), whereas APOBECs are restriction factors against a wide range of retroviruses and retrotransposons (Harris et al. 2003; Mangeat et al. 2003; Arias et al. 2012; Koito and Ikeda 2013). The APOBECs' cardinal role in innate immunity has placed them in a fast paced arms-race against these genomic invaders, resulting in strong positive selection of *apobec* genes (Sawyer et al. 2004; Zhang and Webb 2004; LaRue et al. 2009). This selective pressure has also promoted multiple gene duplications, especially in the placental mammal specific *apobec3* locus, which expanded to a total of seven functional genes in primates (Jarmuz et al. 2002; Conticello et al. 2005; LaRue et al. 2009).

Retrotransposons, which are endogenous retroelements (REs or "elements"), comprise a large fraction of vertebrate genomes (5–10% in typical nonmammalian vertebrates and 30–40% in primates) (supplementary fig. S1 and table S1,

Supplementary Material online). By reverse transcription, the REs multiply and spread throughout the genome. Their dynamism contributes to genomic plasticity and accelerates evolution by altering function of insertion sites, introducing innovative sequences and triggering recombination events (reviewed in Deininger et al. 2003; Kazazian 2004; Feschotte 2008; Cordaux and Batzer 2009). However, they must be restricted to retain genomic stability and avoid detrimental mutagenesis. The APOBECs do so by physically interfering with retrotransposition (Newman et al. 2005; Bishop et al. 2006; Guo et al. 2006; Kaiser and Emerman 2006; Luo et al. 2007; Langlois and Neuberger 2008; Mbisa et al. 2010) and by DNA hyperediting (Harris et al. 2003; Lecossier et al. 2003; Mangeat et al. 2003; Mariani et al. 2003; Esnault et al. 2005; Schumacher et al. 2005; Miyagi et al. 2007; Browne et al. 2009). In the latter, they inflict deleterious hypermutation by inserting a series of C-to-U mutations in retrotransposon single-stranded antisense DNA, right after reverse transcription. The high mutation load that APOBECs cause can impair the retrotransposon's cDNA stability and trigger their degradation (Mariani et al. 2003; Miyagi et al. 2007; Schumacher et al. 2008). In some cases, the retrotransposons can complete mobilization despite being hypermutated, bearing a series of G-to-A mutations in their sense strand after its synthesis based on the C-to-U-edited template (Esnault et al. 2006). Such an

© The Author 2015. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

event results in the insertion of a novel retrotransposon sequence, increasing genomic diversity and the probability of developing a novel functional unit in that genomic locus (Carmi et al. 2011).

Retrotransposons are classified into three major groups: Long terminal repeat retrotransposons (LTRs), LINEs, and SINEs (long- and short-interspersed nuclear elements). LTRs differ from LINEs and SINEs (non-LTRs) in their mode of retrotransposition. LTRs reverse transcribe in a retrovirus-like manner before entering the nucleus (Dewannieux et al. 2004; Esnault et al. 2008), whereas non-LTRs do so by target-primed reverse transcription in the nucleus (Luan et al. 1993; Cost et al. 2002; Dewannieux et al. 2003). In parallel, they are thought to be differentially restricted by APOBECs. The non-LTRs are sufficiently restricted by a deaminase-independent mechanism (Muckenfuss et al. 2006; Stenglein and Harris 2006; Horn et al. 2014), such as interference with reverse transcription, whereas LTR restriction involves or even relies on DNA editing (Dutko et al. 2005; Esnault et al. 2005, 2006, 2008; Bogerd et al. 2006; Dörrschuck et al. 2011). Therefore, in our search for DNA-edited retroelements, we focused on LTRs.

In this study, we computationally screened the genomes of 123 distinct species for edited LTRs. These include human, 11 additional primates, 34 nonprimate mammals, 48 birds, 10 other nonmammalian vertebrates, and 19 invertebrates (supplementary table S1, Supplementary Material online). The LTRs are especially extant in mammals (averaging ~5% of genomic mass) and particularly in primates (6.7%), but comprise smaller portions of nonmammalian vertebrate and invertebrate genomes (typically <4%) (supplementary fig. S2, Supplementary Material online). We expected DNA editing to be vertebrate-specific, as are *apobec* genes, thus the invertebrate genomes were analyzed for negative control.

The APOBECs are well-established retroelement restrictors in mammals. On the contrary, APOBEC activity in nonmammalian vertebrates has not been sufficiently studied. Recently, we identified evidence for retroelement restriction of LINEs by Lizard APOBEC *in vivo* (Lindič et al. 2013). However, concrete support for deaminase-mediated restriction of REs is lacking in nonmammalian vertebrates. Here, we show that DNA editing by APOBECs is not only common in mammals but is exceptionally pronounced in some birds as well. We also provide evidence that hypermutated retroelements created by DNA editing are preferentially retained in active genomic regions, which may pave the way to their accelerated exaptation.

Results

Mammalian and Avian Genomes Contain Many Edited Elements

To date, initial screens found evidence for DNA editing of LTRs in only a handful of organisms, all of which were primates or rodents (Esnault et al. 2005; Carmi et al. 2011; Anwar et al. 2013; Fang et al. 2014), but the scope of editing in other lineages and its impact on genome evolution is yet unknown.

In this study, we screened 80 genomes of different species and various lineages (the UCSC genome browser's reference genomes), for DNA-edited elements and complemented these with a similar analysis of the 43 Avian Phylogenomics Project genomes (<http://avian.genomics.cn/en>, last accessed November 3, 2015) (Zhang et al. 2014), which are addressed in a separate section. In brief, we used Basic Local Alignment Search Tool (BLAST) to pairwise-align REs of the same subfamily in each genome. Then we identified alignments containing clusters of G-to-A mutations in the retroelements' coding strands, the well-known footprint of DNA editing by APOBECs. Such hypermutation suggests that the adenosine-containing element is a DNA-edited version of the guanosine-containing one. For a pair of elements to be declared as such, their alignment had to contain at least ten clustered G-to-A mutations (either one cluster of ten consecutive G-to-A mutations or two of at least five) and pass several validating filters (Materials and Methods). Using a conservative approach that is estimated to generate less than 10% false positive, we identified a total of 6,058 edited elements, containing 137,954 G-to-A edited sites (22.7 editing sites per element). Of note, our method focuses on REs that are hyperedited with high confidence, but the full extent of DNA editing is most probably several-fold greater (supplementary tables S2 and S3, Supplementary Material online).

The edited elements were detected in 32/46 mammalian genomes and 3/15 nonmammalian vertebrates (fig. 1a). In addition to identifying many edited elements in previously unexplored primates (baboon, bushbaby, gibbon, mouselemur, squirrel-monkey, and tarsier) and rodents (guinea-pig and squirrel), we also found edited elements in representatives of a variety of other lineages. These include Marsupials (opossum), Lagomorphs (rabbit), Artiodactyls (cow), Perissodactyls (horse, rhino), Carnivores (ferret, dog), Bats, and more (supplementary data S1, Supplementary Material online). APOBECs from some of these organisms (e.g., rabbit, ferret, horse, cow) have been shown to restrict exogenous retroviruses when overexpressed in *ex vivo* assays (Jónsson et al. 2006; Bogerd et al. 2008; Ikeda et al. 2008) and these results provide evidence that DNA editing by APOBECs plays a role in their genomes *in vivo*.

DNA-edited elements were not limited to mammalian genomes—many exist in avian genomes too (zebra finch, medium ground-finch, and budgerigar). Moreover, the zebra finch exceeded all mammalian genomes in bulk numbers, containing 1,301 edited elements with 35,350 editing sites, which is much more than any other genome tested, and account for 25% of all DNA editing identified. Taking into account the relatively low amounts of LTRs in zebra finch and medium ground finch (42.2 and 11.2 Mb, which are 30% and 8% of the mammalian average, respectively; supplementary table S1, Supplementary Material online), they were enriched in DNA editing 44- and 6-fold more than expected, respectively (Materials and Methods; supplementary table S4, Supplementary Material online). In contrast, applying the same approach to 19 invertebrate genomes shows an absence of DNA editing, as anticipated from genomes lacking the vertebrate-specific *apobec* genes (fig. 1c;

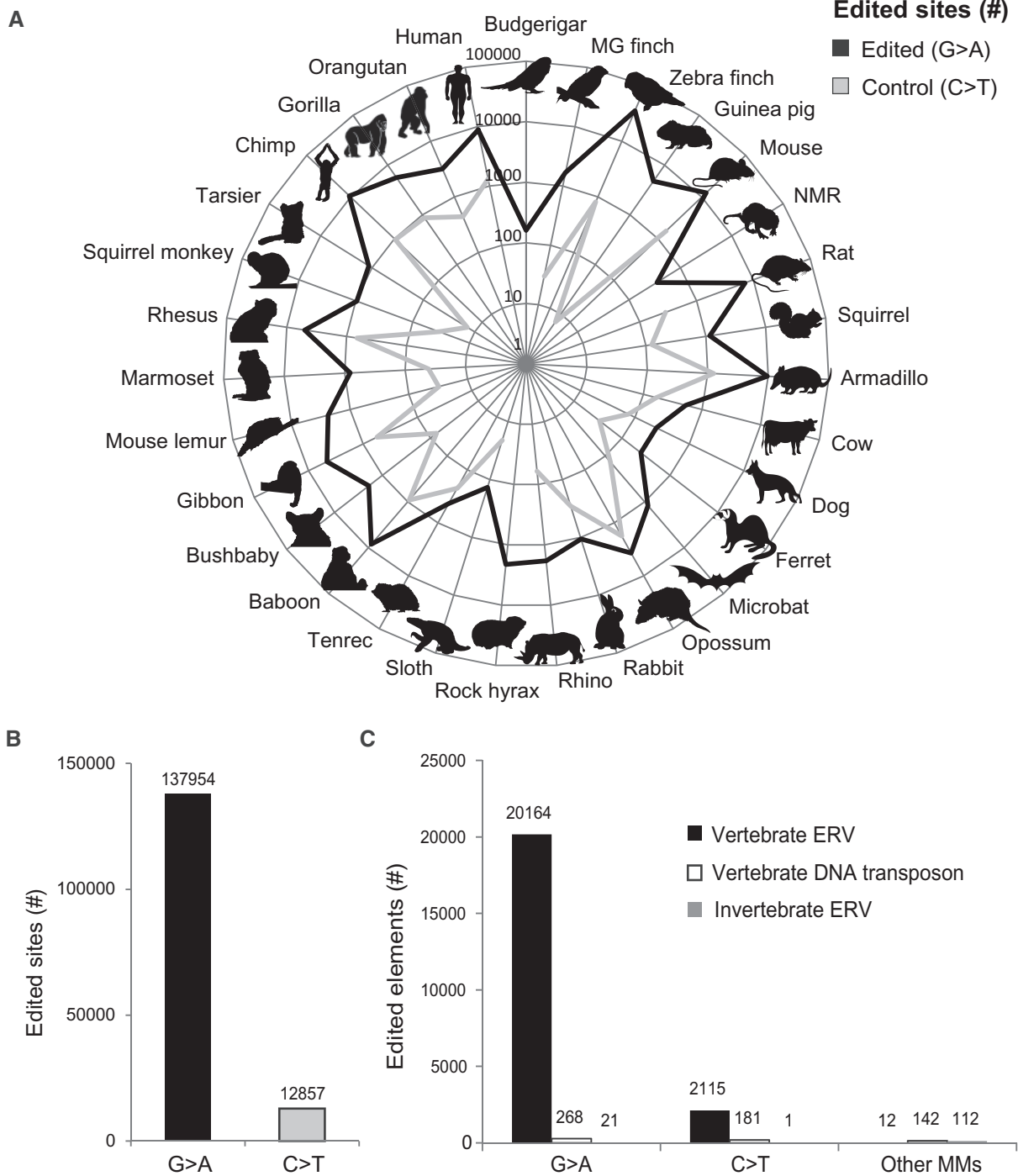


Fig. 1. DNA editing rates in ERVs of vertebrate genomes. (A) The number of G > A DNA editing sites identified in ERVs of vertebrate genomes. As expected from DNA editing by APOBECs, the signal is strongly strand-biased in these genomes, as depicted by the typical 10-fold depletion of the complementary C > T mutations, used as negative control. Only organisms with > 100 edited sites are presented. MG finch, medium ground-finch; NMR, naked mole-rat. The G > A specificity is further pronounced in (B), which sums the results in (A) for all organisms by mismatch type. Mismatches other than G > A and C > T were very rare and therefore were not presented. (C) presents two additional negative controls: 1) Invertebrates are devoid of DNA editing, in concordance with the emergence of APOBECs in vertebrates and 2) DNA transposons, which are not targeted by APOBECs, did not contain signs of DNA editing. Of note, (C) presents results following the first filtering step (the pairwise filter). Analysis of the invertebrates and transversion mutations were not further analyzed, due to their clear depletion at this early phase. MMs, mismatches.

supplementary note S4, Supplementary Material online). *apobec3* genes emerged in placental mammals and are considered the prominent retroviral restrictors in the APOBEC family. Thus, the DNA editing in avian and

marsupial genomes, which is associated with more ancient APOBECs encoded by these genomes, suggests that DNA editing of retroelements is an ancestral function of APOBECs.

Editing Is Strand-Biased and Absent from Invertebrates and DNA Transposons

The well-known signature of APOBEC editing is G-to-A hypermutation of retroelement sense strand DNA (Harris et al. 2003; Lecossier et al. 2003; Mangeat et al. 2003; Mariani et al. 2003; Esnault et al. 2005; Schumacher et al. 2005; Miyagi et al. 2007; Browne et al. 2009). We tested three negative controls to confirm that the G > A hypermutants identified were the product of APOBEC editing and not of random mutagenesis or computational or sequencing artefacts. First, we exploited the strand specificity of APOBEC editing, which causes G > A mutations specifically in the retrotransposon's sense strand, by searching for clusters of the complementary mutation (C > T), which mocks "editing" of the opposite strand. We found that G > A editing rates were over 10-fold greater than C > T transitions (137,954 vs. 12,857 sites, respectively), implying a strong strand-bias, as expected from APOBEC's targets (fig. 1a and b). Furthermore, clustered mutations of all other types were negligible, with a total that is 1,680-fold lower than G > A editing (figs. 1c and 2).

DNA transposons can serve as additional control. These elements transpose in the genome through a "cut-and-paste" mechanism that does not involve reverse transcription. As APOBECs specifically edit single-stranded DNA synthesized during reverse transcription, these elements are not expected to contain strand biased clusters of G-to-A mutations, characteristic of APOBEC editing. Indeed, after the initial filtering step ("pairwise filter," see Materials and Methods) we found only 268 (of 17,912,167) DNA transposons containing G > A clusters, compared with 20,185 (of 21,268,623) LTRs (Fisher's exact; $P < 1e-200$; odds ratio [OR] = 0.015). Moreover, the number of elements bearing other types of clusters was greater than those with G > A, with a total of 323 elements (181 C > T) (figs. 1c and 2). This weak, nonspecific and nonstrand-biased signal implies that the few clusters found in DNA transposons are not associated with APOBECs, unlike the clusters found in LTRs.

The third negative control was a screen for DNA editing in 19 invertebrate genomes. Invertebrates do not encode APOBECs and, therefore, should not contain edited elements. As expected, DNA editing was depleted in invertebrates (Fisher's exact; $P = 5.14e-39$) and was equivalent to noise rates (fig. 1c; supplementary note S4, Supplementary Material online).

Taken together, all three negative controls (strand-bias, DNA transposons, and invertebrate genomes) support that the hypermutation in vertebrate LTRs is indeed APOBEC-mediated DNA editing.

Detection of Known and Novel DNA Editing Motifs

APOBECs have been shown to preferentially edit cytidines in specific sequence contexts (Beale et al. 2004; Liddament et al. 2004). For example, the motif most preferred by mouse APOBEC3 is GxA (underlined G is edited; x is any nucleotide) (Esnault et al. 2005), whereas those preferred by human APOBEC3G and APOBEC3F are GG and GA, respectively

(Beale et al. 2004; Liddament et al. 2004). As expected, the preferences inferred from the mouse and human editing sites identified here resemble their respective APOBEC motifs (Materials and Methods). This concordance suggests that we could infer APOBEC preferences in novel genomes as well and supply a broader understanding of APOBEC editing preferences. As anticipated, we found that the same motifs are present in many related organisms: GxA is the dominant motif in all rodents tested (mouse, Norway rat, naked mole-rat, squirrel, and guinea pig; fig. 3a) and GG and/or GA are the most common preferences in primates (supplementary data S2–S4, Supplementary Material online). Also preferences of uncharacterized APOBECs are similar to known ones: cow and rhino preferences are similar to that of rodents (GxA), whereas finches and armadillo prefer GG like human APOBEC3G (fig. 3d). This is not necessarily expected, because the rapid positive selection of the APOBEC family could have caused their editing preferences to diverge (Sawyer et al. 2004; Zhang and Webb 2004). It would be interesting to understand which selective constraints promoted the motif conservation or convergence in these related and unrelated species. Of note, in addition to the intrinsic APOBEC preferences, DNA repair mechanisms may also act more efficiently on APOBEC-induced G:U mismatches in specific contexts and by doing so contribute their share to the motifs we detected.

The aforementioned preferences represent the most pronounced ones. However, there are secondary preferences that also contribute to the precise motif. By analyzing the nucleotide frequencies adjacent to edited sites of all 35 edited genomes, we learned that the most editing-determining positions are +1, +2 and, to a lesser extent, –1 relative to the edited guanine (see Materials and Methods; fig. 1d; supplementary fig. S3, Supplementary Material online). Based on this observation, these four positions (–1, edited G, +1, and +2) were used to create a profile of 4-mer preferences for each genome and compare editing preferences (Materials and Methods). Hierarchical clustering by these profiles created two major groups—one of primates and another of rodents (fig. 3c and d), both of which contained additional organisms of different taxa. Within primates, the diverged prosimians only weakly clustered with the great apes; and tarsier, which shows a rodent-like profile, was the only outlier. Importantly, the clear separation into two major groups was only evident when clustering by this approach and not when clustering by the retroelement sequences themselves (supplementary figs. S4 and S5, Supplementary Material online). Thus, the clustering effect is attributed to APOBEC motifs and not sequence biases.

For further validation, we wanted to rule out the possible effect of background sequence bias on the detected motifs. The previous analysis was at the organismal level and we wanted to see whether the same motif was present when separately analyzing RE families of a certain genome. To do so we analyzed the medium ground-finch and zebra finch genomes, which had two families with enough editing sites to infer a precise motif. Reassuringly, the motifs were shared by both families within these genomes (fig. 3b). Altogether, the detection of known and recurring preferences supports

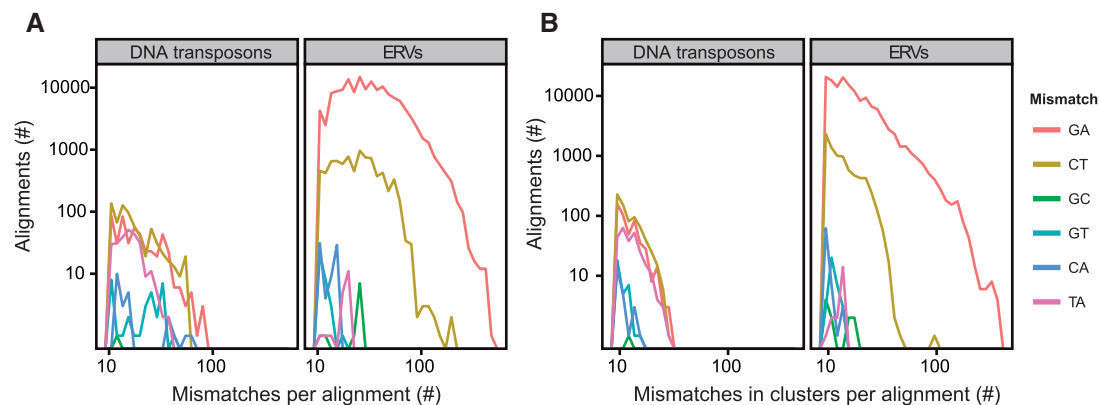


Fig. 2. Prominence of the G > A editing signal in ERVs. Frequency plots portraying the abundance of ERV pairwise alignments containing G > A clusters and their affluence in G > A mutations, in comparison to clusters of other mismatches. The G > A dominance is specific to ERVs (right panels), where G > A DNA editing by APOBECs is expected, and not found in DNA transposons (left panels). Each series shows properties of alignments containing a cluster of a specific mismatch type (G > A, C > T, etc.), where clusters are defined as at least five consecutive mismatches uninterrupted by mismatches of other types (data presented are after the pairwise filter, see Materials and Methods). (A) The number of the specific mismatch in the entire cluster-containing alignment. (B) Similar, but only considering mismatches residing in the clusters.

the validity of the DNA editing sites and supplies a better understanding of APOBEC preferences at large.

Ultra-Edited Elements

The APOBECs' ability to inflict multiple mutations in retroelement sequence enables them to cause long-term impairment of retroelement mobility. In parallel, the heavy mutation load introduced to a relatively short genomic segment accelerates the evolution of that locus, enabling the genome to attain new traits. Consequently, the greater the mutation load, the greater the retroelement inhibition and sequence transformation. Seeking to identify the extremity of APOBEC editing, we searched for "Ultra-edited" (UE) elements. We selected elements with at least 25 G-to-A mutations and demanded that they comprise the majority of all mismatches in the alignment to their tentative unedited parent element. This analysis revealed 27 genomes containing UE elements, with a total of 1,237 such elements (see fig. 4 for examples; supplementary table S5, Supplementary Material online). The organism with the highest number of UE elements was zebra finch (570), followed by mouse (109), chimp (80), baboon (73), and human (44). There was a correlation between the amount of UE and edited elements per organism (Pearson's $r = 0.858$, $P = 4.283e-11$), implying that ultra-editing is a common outcome of DNA editing by APOBECs.

The most edited element is a member of the zebra finch ERVL family, containing 353 G-to-A editing sites (fig. 4a; 392 G-to-A transitions in an alignment 4,304 bp long, with an estimate of only 39 mutations not associated with APOBEC; Materials and Methods). The zebra finch ERVK family also contains an extremely UE element with 332 edited nucleotides (353 G-to-A and 21 background). To the best of our knowledge, these are mutation loads unprecedented by findings in any other genomic analysis of retrotransposons. Primates and rodents also contain extensively UE elements, with up to 234 editing sites, which would have taken over 100 million years of evolution to occur by random mutagenesis (fig. 4; Materials and Methods; supplementary

data S5, Supplementary Material online). We also identified an UE mouse intracisternal A particle (IAP) retroelement, bearing 189 edited sites. This was anticipated, as IAP elements have been shown to be edited by murine APOBEC3 (Esnault et al. 2005; Carmi et al. 2011) (fig. 4).

We analyzed the motifs of editing sites in the most UE element in each family per genome. Reassuringly, the motifs most commonly resembled those inferred from all editing sites in the respective genomes (e.g., $\underline{G}xA$ in rodents and $\underline{G}A$ or $\underline{G}G$ in primates; supplementary data S5, Supplementary Material online). We also ruled out the possibility that UE elements were artefacts of sequencing errors (supplementary note S6, Supplementary Material online) (Zarnek et al. 2010). In summary, the UE elements are high-confidence DNA editing examples that portray the extremity of this process and demonstrate that DNA editing can profoundly transform retroelement sequences.

The Full Scope of DNA Editing Is Yet Unmasked

Upon insertion of an edited element, it is virtually identical to its progenitor element, except for the G-to-A editing sites. Over time, random mutations accumulate in both elements and mask the editing signal, making editing detection in their sequence alignment hard or even impossible. Therefore, we expected the edited elements to be enriched in relatively recent or "young" insertions. To categorize the elements by time of insertion, we checked which were species-specific and which were present in genomes of related species, implying insertion in a common ancestor. Testing this within hominids, rodents and songbirds revealed that the edited elements in all three lineages were enriched in species-specific elements, as expected (chi-square P values 3.33E-37, 1.19E-38, and 1.74E-22, respectively; ORs 2.69, 23.28, and 2.39, respectively; Materials and Methods). The enrichment in species-specific elements is even greater in UE elements, to the extent that all 148 UE elements in rodents were species-specific (chi-square P values 2.59E-19, 1.39E-07, and 9.63E-10 for the three lineages, respectively, when compared with other edited elements).

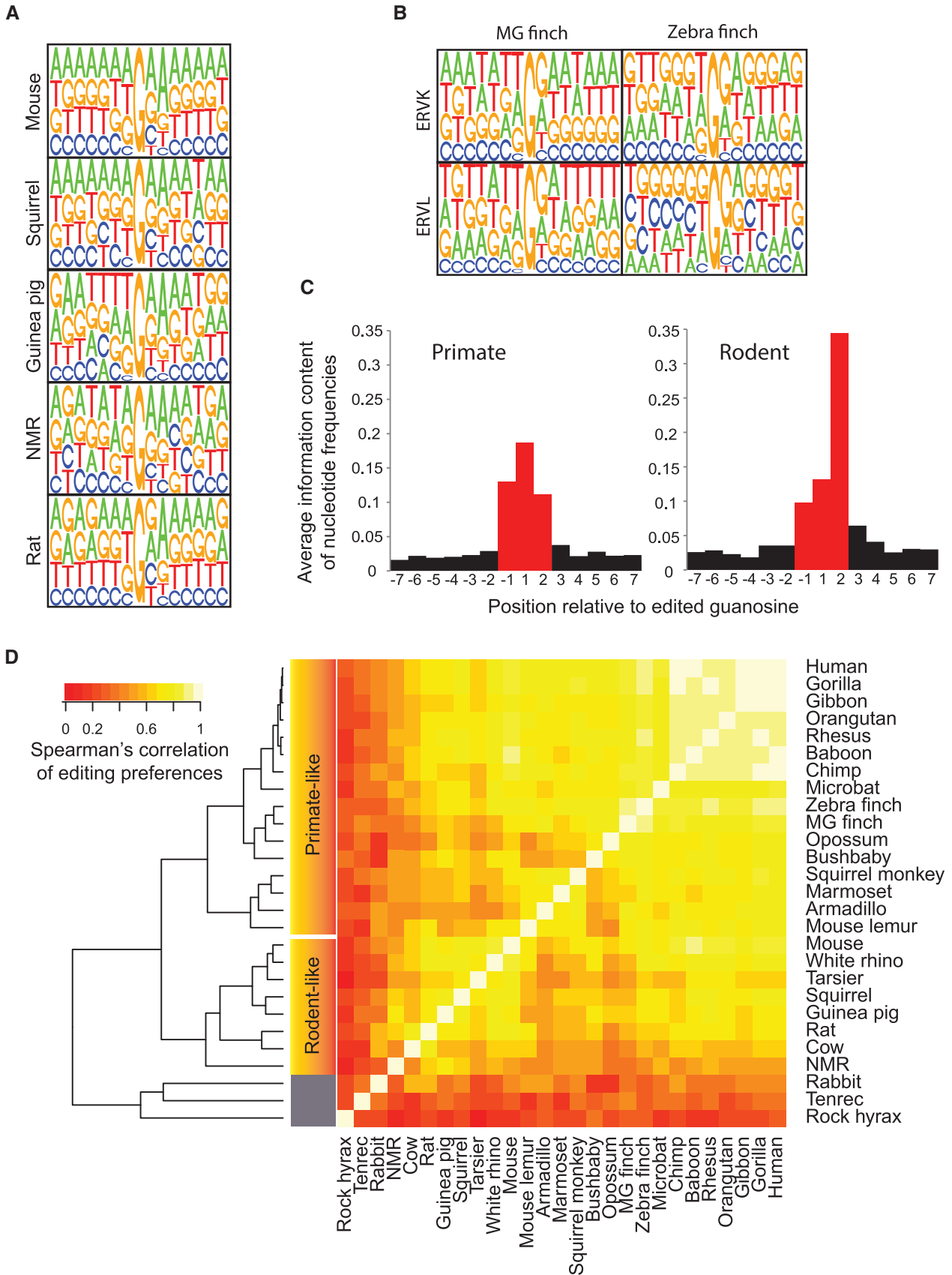


FIG. 3. DNA editing preferences of mammalian and avian APOBECs. We inferred the editing preferences of APOBECs in each genome from the editing sites identified. (A) shows that all rodents strongly prefer A at+2 downstream, as known for mouse APOBEC3. (B) Songbirds share a strong preference for G at+1, which was independently inferred from two families of ERVs in each genome. These consistent motifs validate both the editing sites identified and the motifs themselves. (C) The dependency of DNA editing on positions adjacent to the edited G (see Materials and Methods).

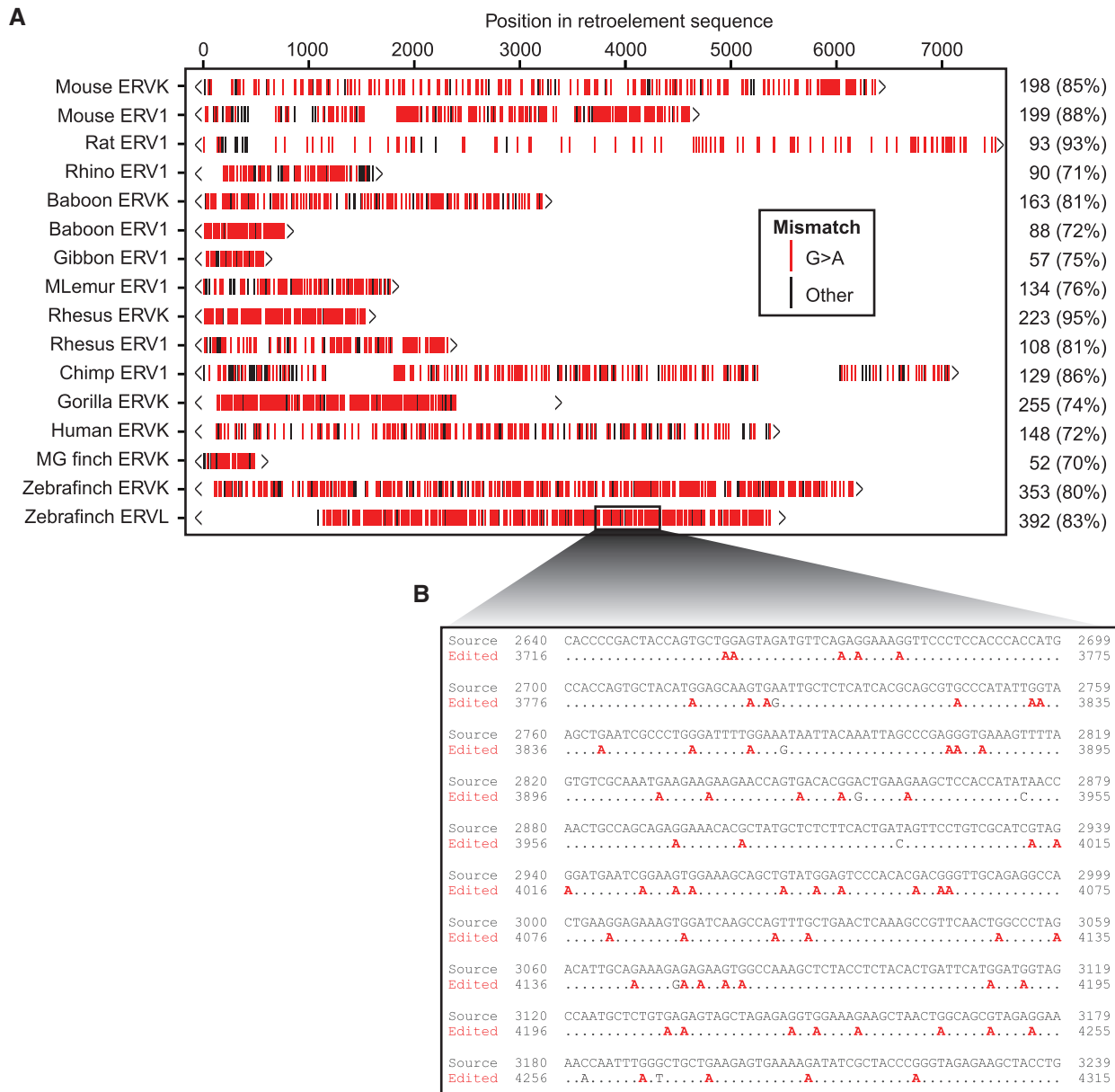


Fig. 4. Ultra-edited ERVs in vertebrate genomes. (A) Examples of pairwise alignments between the most hyperedited elements and unedited elements within various genomes, as produced by BLAST in the computational pipeline. Every row resembles an element from a different retroelement family, where red bars are G > A mutations and black bars are any other mutation type. The values on the right are the amount of G > A mismatches and their percentage of all mismatches in the respective alignment (full alignments are available in [supplementary material, Supplementary Material](#) online). These mutation loads are extremely high and would take tens of millions of years to accumulate without DNA editing. The “<” and “>” signs indicate the ends of the retroelement sequence. Long stretches without mismatches in the chimp, gorilla, and zebra finch ERVL alignments are regions that were not aligned by BLAST. (B) Blowup of a region in the BLAST alignment between an edited zebra finch ERVL element and its tentative unedited source element. The full alignment contains a total of 392 G > A mutations (83% of all mismatches in the alignment). In the edited element, dots represent matches, G > A mismatches are marked in red, and all other mismatches are black letters.

Fig. 3. Continued

Dependencies for organisms with primate-like and rodent-like motifs are presented in the left and right panels, respectively. Both groups are most dependent on -1, +1, and +2. The strong dependency of rodent APOBEC on the +2 position is clear, whereas primate APOBECs depend more on +1. The lower signal in primate +1 is due to differential preferences of distinct APOBECs encoded in primate genomes, preferring either G or A at +1. (D) Hierarchical clustering of organisms by editing preferences. After identifying that the -1, +1 and +2 positions are most editing-determining we calculated frequencies of 4-mers (comprising these three positions and the edited G) of all editing sites in each genome (see Materials and Methods). Clustering and heatmap intensity are based on Spearman correlations of 4-mer frequencies between genomes. Related organisms tend to have similar preferences. Additionally, two major clusters form, containing organisms with rodent- or primate-like preferences, which seem to reflect two archetypical APOBEC preferences. These two groups were those used for the separate panels in (C).

This suggests that many edited and UE elements have accumulated mutations and are currently excluded from the high confidence cohorts we put forth.

To further support this finding, we hypothesized that genomes more abundant in young and intact retroelements will be rich in editing, due to ease of detection in such sequences. As expected, there was a positive correlation between the amount of such retroelement sequences and the abundance of DNA editing in a given genome ($r = 0.827$, $P = 2.12E-06$; [supplementary fig. S6, Supplementary Material online](#); Materials and Methods). Finally, we wanted to see whether the various filtering steps, applied for high confidence, discard potential editing-containing elements. Analyzing the intermediate sets of tentative edited elements revealed that these larger cohorts already contain the DNA editing motif identified after the final filtering step (tested for human, mouse, and zebra finch; [supplementary fig. S7, Supplementary Material online](#)). Put together, these independent lines of evidence support that the prevalence of DNA editing is greater than currently quantified.

DNA Editing Plays a Role in Genome Defense in Diverse Lineages

APOBECs are important players in antiviral innate immunity in humans (Refsland and Harris 2013). However, their importance as retroelement restrictors has been unexplored in most organisms. The affluence of DNA-edited retroelements we identified suggests that APOBECs restrict retroelements in various vertebrate lineages. To support the notion that DNA editing helps attenuate the edited retroelements, we analyzed the edited sites in the retroelement sequences. Indeed, many editing sites reside in annotated retroelement open-reading frames (ORFs) (Materials and Methods; [supplementary note S7 and table S6, Supplementary Material online](#)). To verify that the mutations within ORFs impair the retroelements' coding capacity, we characterized the effect of editing-induced mutations on zebra finch ERVL ORFs. Of the 1,761 editing sites analyzed, we identified 225 mutations that created premature stop-codons by altering at least one of the guanines in TGG tryptophan codons. Additionally, editing produced an abundance of nonsynonymous mutations (989), > 2-fold more than synonymous ones (449), which may compromise protein function ([supplementary fig. S8, Supplementary Material online](#)). This abundance of ORF-compromising mutations in an avian genome shows that retroelement restriction through DNA editing is not limited to APOBEC3-encoding placental mammals.

The *apobec*s have been under strong positive selection throughout mammalian evolution (Sawyer et al. 2004; Zhang and Webb 2004). Despite being famous for their arms race with HIV, the *apobec* family expansion predates lentiviruses, and is probably associated with their role in restricting retroelements (Sawyer et al. 2004; Zhang and Webb 2004; Refsland and Harris 2013). Therefore, we wanted to see which ERV families have been involved in this interaction. We hypothesize that the relative abundance

of edited elements in an ERV family correlates with their extent of collision with APOBECs. We found that the family containing the most edited sites is ERV1 (69,495 sites; [fig. 5](#)), but ERVK and ERVL are more enriched in editing ([supplementary table S7, Supplementary Material online](#)). This discrepancy is caused mostly by the skyscraping rates of editing in songbird ERVK and ERVL. The predominance of ERVK editing is also evident in primates and rodents, but the remainder of editing is present in ERV1 and not ERVL ([fig. 5](#)). The most enriched family is bird ERVK, where approximately 1 in every 1,000 bases is edited ([fig. 5; supplementary data S1, Supplementary Material online](#)). Further scrutinizing editing per subfamily revealed that a total of 473 distinct subfamilies were edited in the analyzed genomes ([supplementary table S8 and data S6, Supplementary Material online](#)). In concordance with the edited site per family analysis, ERVK and ERV1 contained the greatest number of edited subfamilies (202 and 188, respectively). The fact that DNA editing is not limited to a small subset of subfamilies implies that APOBECs provide their hosts with widespread defense against retroelements ([supplementary fig. S9, Supplementary Material online](#)).

Edited Elements Are Enriched in Transcriptionally Active Genomic Regions

Retroelements contribute to genome dynamism and accelerate its evolution (reviewed in Deininger et al. 2003; Kazazian 2004; Feschotte 2008; Cordaux and Batzer 2009). It has been proposed that DNA editing of retroelements adds another layer of flexibility for genomic innovation (Carmi et al. 2011). To understand the impact of DNA-edited elements on their host genomes, we characterized the genomic location of edited elements and sites. We identified 1,019 edited elements intersecting with genes (Materials and Methods). Of these, the vast majority (1,000) overlap introns and 45 overlap exons, including 8 protein-coding ones ([supplementary tables S9 and S10, Supplementary Material online](#)). In comparison to unedited elements in the same subfamilies, the edited elements are not only enriched in genes (chi-square test; $FDR = 1.96E-05$; $OR = 1.17$), but are also preferentially exonized ($FDR = 2.25E-06$; $OR = 2.09$; [fig. 6](#)). This suggests that DNA editing can increase a retroelement's probability of being retained in genic regions and hence their potential to be exapted. In line with this, by analyzing data from human cell lines (Materials and Methods), we found that edited elements are also enriched in 1) histone modifications associated with active chromatin (especially H3k4me2; [supplementary table S11, Supplementary Material online](#)), 2) DNase hypersensitive regions (Fisher's exact $P = 2.44E-15$, $OR = 3.341$), and 3) transcription factor-bound DNA (Fisher's exact $P = 2.44E-15$, $OR = 2.691$; [supplementary figs. S10–S12, Supplementary Material online](#)). The edited elements are not enriched in any specific biological pathway or gene ontology (GO) term, as expected from a general mutation mechanism ([supplementary note S11, Supplementary Material online](#)). Interestingly, the strongest enrichment in exonization of edited elements is in the human genome, having a greater than 4-fold enrichment over unedited elements (26 elements; Fisher's exact;

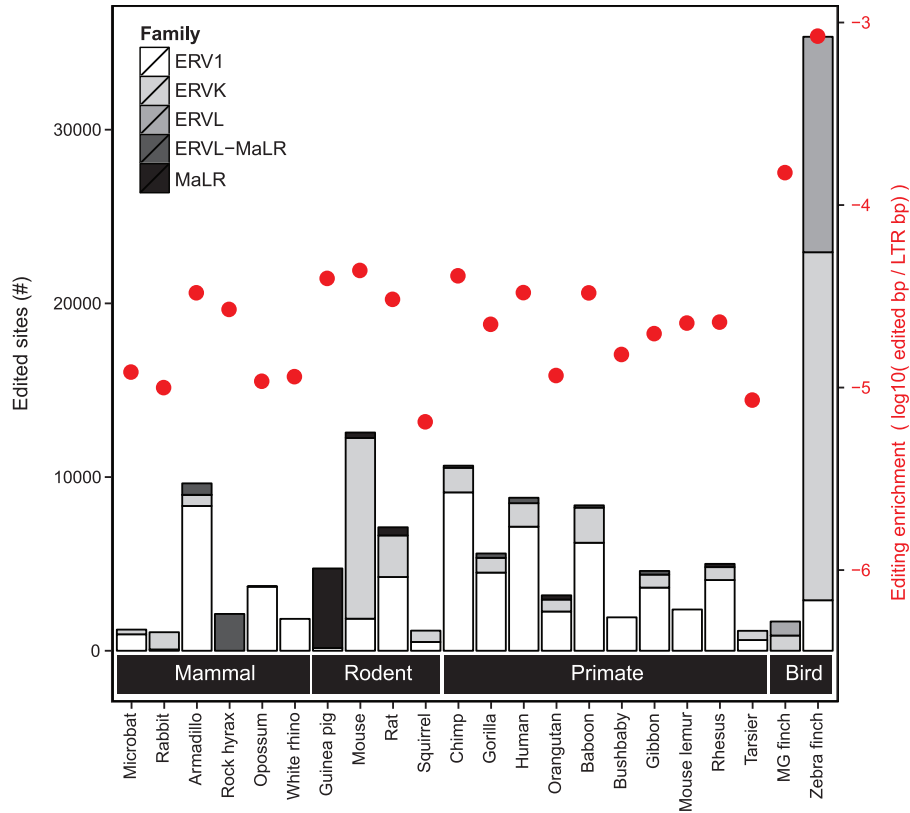


Fig. 5. DNA editing rates in genomes most abundant in editing. Only genomes with > 1,000 edited sites are presented. Columns show amount of DNA editing and its distribution between ERV families per genome (left axis). The dots depict enrichment of DNA editing in each organism, based on total LTR retrotransposon content of each genome (right axis).

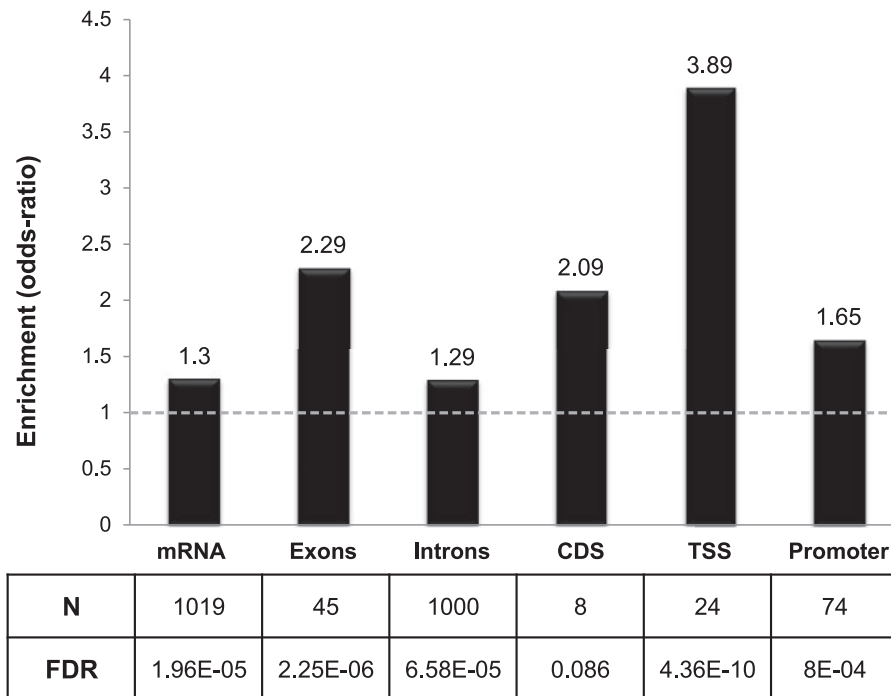


Fig. 6. Edited elements are enriched in active genomic regions. Comparing edited with unedited retroelements shows that the edited ones are enriched in expressed and/or functional genomic regions (in the cohort of genomes containing edited elements). The columns present ORs of the fraction of edited elements residing in a genomic region in comparison to the fraction of unedited elements residing in it. *N*, number of retroelement sequences overlapping each genomic region. FDR, *P* values from chi-squared test corrected for multiple testing, depicting the significance of enrichment in each region. Promoters were defined as regions within 1,000 bp upstream of a TSS.

FDR = 6.56E-13). Other organisms contained edited elements in genes and enrichment was observed in the guinea pig and zebra finch genomes (FDR < 0.05; [supplementary table S12, Supplementary Material online](#)).

Editing is not necessarily the last step toward exaptation of a retroelement. An edited element may need to undergo additional mutations before it is utilized by the genome. Thus, we hypothesize that older edited elements, which have accumulated more mutations over time, have a greater probability to be exapted. Indeed, comparing ancient edited elements (i.e., inserted in the hominids' common ancestor) with younger ones shows that the ancient ones are overrepresented in genic regions (Fisher's exact $P = 0.00044$; OR = 2.98) and have been preferentially exonized (11/12 exonized elements; > 5-fold increase) in the human genome. Of note, this effect is associated with editing and not solely with the age of the element, as even within the set of ancient elements, the edited ones were more likely to reside within genes and exons than the unedited ones (Fisher's exact; $P = 0.017$, OR = 1.5 and $P = 0.0002$, OR = 4, respectively). Nonetheless, in the whole set of retroelements analyzed, some of the enrichment in genes may be associated with the enrichment of edited elements in young elements, which have not yet had time to be sifted out of the genome.

Genomic LTRs can regulate transcription of neighboring genes (Wang et al. 2007). We searched for signs of such regulation mediated by edited elements and found that edited elements are enriched in transcription start sites (TSSs) and promoter regions (Fisher's exact; $P = 1.14E-05$, OR = 3.17 and $P = 0.049$, OR = 1.29, respectively; [supplementary tables S13 and S14, Supplementary Material online](#)). These events were most apparent in the human genome (13/20 TSSs; 18/62 promoters), where enrichment was even stronger ($P = 6.65E-08$, OR = 7.43 and $P = 6.24E-4$, OR = 2.51). Some of these elements contain TSSs of novel coding and noncoding mRNA isoforms ([supplementary fig. S12, Supplementary Material online](#)). In addition to these alterations in transcriptional regulatory regions, edited elements can also be utilized to create splice variants. For example, we identified an edited element residing in the mouse *Ifi203* gene that gives rise to a novel exon by alternative splicing (both donor and acceptor splice sites are contained in the edited element; [supplementary fig. S13, Supplementary Material online](#)).

DNA Editing Exists in a Variety of Avian Clades

The extensive editing in the zebra finch genome prompted us to seek a wider picture of DNA editing in avian genomes. To do so, we analyzed 43 avian genomes from the recently published Avian Phylogenomics Project (Zhang et al. 2014) and identified DNA editing in 34 of them ([supplementary tables S15 and S16, Supplementary Material online](#)). In total, these genomes contained 9,914 edited sites (and 9-fold more when relaxed; [supplementary table S17, Supplementary Material online](#)), which was > 5.5-fold more than control C-to-T mutations (1,777). Interestingly, of all 123 genomes analyzed, the 5 organisms most enriched in DNA editing

following zebra finch and medium ground finch were also birds—kea, white-tailed tropicbird, northern fulmar, carmine bee-eater, and adelic penguin—all of which are not songbirds ([supplementary table S15, Supplementary Material online](#)). Among the 3 songbirds in these 43 genomes (crow, rifleman, and golden-collared manakin), only the crow was enriched (1.24-fold) in DNA editing when compared with the other birds, yet all 3 bore a motif resembling that of zebra finch and medium ground-finch (GG). The bird with the most heavily edited element was the carmine bee-eater, containing two elements with greater than 150 G-to-A edited sites. Thus, DNA editing seems to serve in retroelement restriction in a wide range of avian species.

Discussion

DNA Editing Is Widespread in Mammals and Birds

In this study, we found clustered mutations implying that APOBECs edit DNA of LTRs in genomes of 69 species in various vertebrate lineages. As expected, many placental mammals (32), which encode the potent antiviral protein APOBEC3, contain signs of DNA editing. However, we supply first solid evidence that APOBECs restrict ERVs in nonplacental genomes too, namely, marsupials (2) and birds (37). Intriguingly, not only is DNA editing evident in these genomes, but zebra finch also displays an unprecedented abundance of editing. Complementing these results with the analysis of the Avian Phylogenomics Project genomes revealed that DNA editing is common among songbirds and evident in most avian genomes analyzed. This cohort of avian genomes contains representatives of all major avian clades, enabling us to conclude that APOBEC activity is widespread in birds.

APOBEC3 proteins, which emerged in placental mammals, were once thought to be the predominant ERV restrictors. However, Ikeda et al. (2008, 2011) showed that APOBEC1 of some placental mammals (rabbit and rodents) potentially restrict retroelements by DNA editing. This and the recent finding that lizard APOBEC1 can restrict retroelements in *ex vivo* assays (Lindič et al. 2013) suggest that APOBEC1 may play this role in non-placentals. Interestingly, APOBEC1 was duplicated in amniotes, similar to APOBEC3 in placentals (Severi et al. 2011). It is possible that an analogous selective pressure by retroelements caused both duplications. Another, yet less supported, candidate is AID, which has been shown to weakly restrict the ERV *MusD* (MacDuff et al. 2009). APOBEC5, whose function is unknown, is an APOBEC homolog most similar to APOBEC1 and APOBEC3 encoded by marsupials and some amniotes (Severi et al. 2011). It may be responsible, at least in part, for DNA editing in these genomes. We expect future experimental studies to unravel which APOBECs are the DNA editors in each lineage.

DNA Editing Diversifies Retrotransposon Sequence

Our study shows that DNA editing creates diversity in retrotransposon populations. We identified over 6,800 hyperedited elements, containing more than 20 edited

sites on average. Such hyperedited sequences are instantly transformed during retrotransposition, generating unique retrotransposon sequences. One immediate implication is that DNA editing should be taken into account when assessing retrotransposon age, commonly inferred by divergence from an ancestral sequence (Smit 1999). Importantly, there are several reasons to believe that the effect of DNA editing is much greater than presented here. First and foremost, the approach we used is conservative and tends to identify recently hyperedited elements. However, *ex vivo* studies show that there is high variance in the number of edited sites per retroelement, including many elements (possibly the majority) with only few mutations (Esnault et al. 2005, 2006; Suspène et al. 2011). Thus, vertebrate genomes probably contain many moderately edited elements yet to be identified. Indeed, relaxing the algorithm's parameters reveals many additional edited elements, albeit with lower confidence of editing (supplementary table S2, Supplementary Material online). Second, we focused on homogenous G-to-A mutations. However, AID/APOBEC-induced G:U mismatches can be altered by various DNA repair mechanisms, leading to heterogeneous clusters of mutations. Third, retroelement annotation of many of the genomes analyzed is incomplete, due to insufficient consensus sequence libraries needed for identification by RepeatMasker (Lindič et al. 2013). Advancement in genomic annotation will help understand the full extent of DNA editing in these genomes. Thus, the diversity that DNA editing causes in vertebrates is probably much greater than currently observed.

DNA Editing Can Accelerate Genome Evolution

We found that edited elements are overrepresented in genes, exons, promoters, and TSSs. Intriguingly, this is in stark contrast to the typical underrepresentation of LTRs in human genes, which has been associated with purifying selection (Smit 1999; Nellåker et al. 2012). A possible explanation for the preferential retention of edited elements in genes is that edited elements, which bear high mutation loads, impose less of a threat on the genome, hence are less selected against. Additionally, retroelements may be preferentially retained due to positive selection (Tsirigos and Rigoutsos 2009). Therefore, it is possible that in a few cases, the accelerated sequence transformation caused by DNA editing, which introduces many mutations in a single locus in one generation, gives rise to a beneficial sequence that is positively selected.

Conclusions

In conclusion, our screen for DNA editing of retrotransposons in diverse genomes sheds light on the role of APOBECs in vertebrate genomes. We found that DNA editing of retrotransposons is evident in marsupials, most common in placentals and at highest rates in some birds, such as zebra finch. Ultimately, we show that the DNA editing has had an important role in deactivating retrotransposons and may have accelerated genome evolution in a wide range of vertebrates.

Materials and Methods

Data

Eighty animal genomes accessible through the UCSC genome browser (Karolchik 2003) were analyzed (supplementary table S1, Supplementary Material online). Genomic sequences of LTR retrotransposon and DNA transposon of all assemblies were downloaded using Galaxy (Giardine et al. 2005), based on the genomic coordinates found in the UCSC table browser's (Karolchik et al. 2004) RepeatMasker (Smit et al. 1996) table.

Detection of DNA-Edited Retroelements

DNA editing of retroelements by APOBECs introduces G-to-A mutations specifically in the retroelements' sense strand. The mutations are nonuniformly distributed throughout the sequence and tend to be found in clusters. In our approach, we assume that edited elements have unedited "ancestral" elements in the genome. Thus, an edited element should be very similar to one or more ancestral or "source" elements, except for the editing sites, where the source will contain guanines and the edited element adenosines. To detect clustered mutations, we aligned pairs of LTRs of the same subfamily in each genome using BLAST v2.2.23 (for parameters and various details regarding the entire method, see supplementary note S1, Supplementary Material online). Then, we identified clusters of ≥ 5 consecutive G-to-A mismatches (i.e., uninterrupted by any other mismatch) and chose alignments containing at least 10 clustered G-to-A mutations in total.

Next, we applied two filters to select alignments most likely containing APOBEC-associated G-to-A mutations. We refer to these as the "pairwise filter" and "consensus filter." In the former, we selected only alignments with dominance of unidirectional (G-to-A and not A-to-G) and strand biased (G-to-A and not C-to-T) mutations (supplementary table S20, Supplementary Material online). The consensus filter uses the consensus sequence of each element's subfamily to identify which of the two aligned sequences is most probably the ancestral one. This information is crucial to ascertain that the direction of the mutations is G-to-A and not the opposite. In detail, we demanded that 1) most of G-to-A editing sites read G in the consensus (by aligning the edited element to the consensus using BLAST) and 2) the adenosine-containing element be more diverged from the consensus than the guanine-containing one. Together, these constraints imply that the edited element diverged from the consensus by G-to-A DNA editing.

Assessment of Edited Sites per Element

Our screen for "clustered" G-to-A mutations effectively reduces false-positives. However, in its conservative nature it underestimates the amount of edited sites per retroelement. To assess the total number of edited sites in each element (not only those in clusters), we counted all G-to-A mismatches in the alignment to the element's tentative source element. From this value we subtracted the second most common mismatch, which estimates the mutation rate caused by

random mutagenesis since the edited element was inserted. To identify the most probable source element per edited element, we first identified all elements whose alignment with the edited element exhibited a signal of DNA editing (as defined above) and then chose the element that generated the alignment with the highest BLAST bitscore.

Negative Controls

1) The strand-biased nature of APOBEC DNA editing (with regard to the retroelement's ORFs) enables us to use C-to-T, the event complementary to G-to-A, as control. So we searched for clusters of C-to-T mutations, using the same data and algorithm. 2) DNA transposons are seemingly not targeted by APOBECs, thus should not contain strand-biased mutations. We searched for all types of clustered mutations in DNA transposons of the 80 genomes to validate this. 3) Invertebrates do not encode APOBECs. Thus, the 19 invertebrate genomes analyzed were expected to lack signs of DNA editing.

Enrichment of DNA editing per organism was calculated using ORs and chi-square test in R, comparing the number of editing sites per LTR base pair in the genome for each genome against the other genomes that had edited elements.

Estimation of evolutionary time needed to generate mutation rates similar to those in edited elements was calculated by the number of editing sites per the element's alignment length, using a mutation rate of 1×10^{-9} mutbp⁻¹*yr⁻¹ (as in Prado-Martinez et al. 2013).

Editing Preference Inference

To infer editing preferences, we first identified the most editing-determining positions in each retroelement family of every genome. We were interested in positions relative to the edited guanosine whose nucleotide frequencies significantly differed from background frequencies. The latter were inferred from nucleotide frequencies proximal to all guanosines in the retroelement family. Divergence from background was tested by chi-square test for each position and then FDR adjusted. To quantify divergence from background, Shannon's information content (IC) was calculated for the nucleotide frequencies in each position, using R's seqLogo package (v1.28.0) (Bembom 2007). IC for positions with FDR > 0.05 was set to 0. Average IC of all families was calculated for each position and those with markedly elevated values were designated most editing-determining. In our case, these were -1, +1, and +2 relative to the edited G (supplementary fig. S3, Supplementary Material online).

We used these three dominant positions to see whether related organisms prefer the same motifs in these positions. So we calculated the relative frequency of each 4-mer (i.e., -1, edited G, +1, and +2) per organism, creating a full profile of 4-mer preferences for each organism. Of note, only organisms with ≥200 editing sites were analyzed, to ensure high confidence of preferences. Then, we calculated Spearman correlations of profiles between organisms. Finally, we used these

correlations to cluster organisms together using the *heatmap.2* function in R for hierarchical clustering by Pearson correlation of the Spearman correlation vectors, with default clustering parameters.

In addition to these analyses at the familial and organismal levels, we identified editing motifs in individual retroelements, as previously described (Carmi et al. 2011). We used this method to identify the most common motifs per organism and motifs of the UE elements.

Genomic Location of Edited Elements

Ensembl (Hubbard et al. 2002) gene tables were downloaded from the UCSC table browser's FTP site (<ftp://hgdownload.cse.ucsc.edu/goldenPath>) for all assemblies (where available; supplementary table S19, Supplementary Material online). Coordinates of genomic regions (genes, exons, introns, promoters, and TSSs) were extracted using PERL scripts and were intersected with edited elements. Promoters were defined as regions within 1,000 bp upstream of TSSs. Enrichment of edited elements in genomic regions was tested versus unedited elements by chi-square test and FDR in R. To reduce potential biases, the set of unedited elements used as background included only those belonging to subfamilies containing edited elements. Additionally, subfamilies not intersecting any of the genomic tracks were not included in the statistical tests.

Similarly, edited elements in human were tested for enrichment of histone modifications, DNase hypersensitivity, and TF binding. All data files analyzed were part of the ENCODE project and downloaded from the UCSC genome browser's FTP site. For histone modifications, H1-hESC and HUVEC data were analyzed from the broad histone modification files. <wgEncodeRegTfbsClusteredWithCellsV3.bed.gz> was analyzed for TF binding. <wgEncodeRegDnaseClusteredV2.bed.gz> was analyzed for DNase hypersensitivity; Enrichment in GO terms or biological pathways of human genes containing edited elements was tested using STRING (Franceschini et al. 2013).

Species-Specific Enrichment and Evolutionary Branch Analyses

Retroelements of interest were assessed for time of insertion by checking for their presence in syntenic regions of related organisms. This included intralinea analyses in hominids (human, chimp, gorilla, and orangutan), rodents (mouse, rat, and guinea pig), and songbirds (medium ground-finch and zebra finch). Edited and unedited elements (of subfamilies containing edited elements) of each organism were tested for presence in the other genomes of the same lineage by using the executable of UCSC genome browser's Liftover tool (June 2013). Elements were designated either species-specific, if absent from all other assemblies (i.e., not fully or partially lifted), or not so, if present in entirety in any other organism. Enrichment of edited elements in species-specific elements was tested by chi-square test in R.

Correlation between Young LTR Content and Edited Site Count

Basepair count of LTR elements $\leq 10\%$ divergence (mismatches) from consensus and $\geq 1,000$ bp in length were summed per genome. Genomes having less than 10^5 bp of such LTR element sequence were excluded from analysis to reduce biases, retaining 24 of 35 editing-containing UCSC reference genomes. Spearman's correlation was applied to final numbers of editing sites per genome.

Avian Phylogenomics Project Analysis

Forty-three Avian genomes were downloaded from the Avian Phylogenomics Project database (<http://avian.genomics.cn/en/jsp/database.shtml>, last accessed November 3, 2015). The same approach for DNA editing identification used for the 80 UCSC reference genomes was applied (with the exception that whole families and not only subfamilies were pairwise aligned, for technical reasons). Enrichment per organism was calculated similarly.

Supplementary Material

Supplementary notes S1–S12, figures S1–S16, tables S1–S23, data S1–S8, and supplementary alignments are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors thank Shai Carmi for his advice and contribution of PERL scripts. They also thank Nika Lovšin for her helpful discussions. This work was supported by the European Research Council (grant no. 311257), the I-CORE Program of the Planning and Budgeting Committee in Israel (grant nos 41/11 and 1796/12), and the Israel Science Foundation (1380/14).

References

Anwar F, Davenport MP, Ebrahimi D. 2013. Footprint of APOBEC3 on the genome of human retroelements. *J Virol.* 87:8195–8204.

Arias JF, Koyama T, Kinomoto M, Tokunaga K. 2012. Retroelements versus APOBEC3 family members: no great escape from the magnificent seven. *Front Microbiol.* 3:275.

Beale RCL, Petersen-Mahrt SK, Watt IN, Harris RS, Rada C, Neuberger MS. 2004. Comparison of the differential context-dependence of DNA deamination by APOBEC enzymes: correlation with mutation spectra in vivo. *J Mol Biol.* 337:585–596.

Bembom O. 2007. seqLogo: an R package for plotting DNA sequence logos. Available from: <http://www.bioconductor.org/packages/release/bioc/html/seqLogo.html>.

Bishop KN, Holmes RK, Malim MH. 2006. Antiviral potency of APOBEC proteins does not correlate with cytidine deamination. *J Virol.* 80:8450–8458.

Bogerd HP, Tallmadge RL, Oaks JL, Carpenter S, Cullen BR. 2008. Equine infectious anemia virus resists the antiretroviral activity of equine APOBEC3 proteins through a packaging-independent mechanism. *J Virol.* 82:11889–11901.

Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. 2006. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. *Nucleic Acids Res.* 34:89–95.

Browne EP, Allers C, Landau NR. 2009. Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. *Virology* 387:313–321.

Carmi S, Church GM, Levanon EY. 2011. Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. *Nat Commun.* 2:519.

Corticello SG, Langlois M-A, Yang Z, Neuberger MS. 2007. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv Immunol.* 94:37–73.

Corticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Mol Biol Evol.* 22:367–377.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 10:691–703.

Cost GJ, Feng Q, Jacquier A, Boeke JD. 2002. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* 21:5899–5910.

Deininger PL, Moran JV, Batzer MA, Kazazian HH. 2003. Mobile elements and mammalian genome evolution. *Curr Opin Genet Dev.* 13:651–658.

Dewannieux M, Dupressoir A, Harper F, Pierron G, Heidmann T. 2004. Identification of autonomous IAP LTR retrotransposons mobile in mammalian cells. *Nat Genet.* 36:534–539.

Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet.* 35:41–48.

Dörrschuck E, Fischer N, Bravo IG, Hanschmann K-M, Kuiper H, Spötter A, Möller R, Cichutek K, Münk C, Tönjes RR. 2011. Restriction of porcine endogenous retrovirus by porcine APOBEC3 cytidine deaminases. *J Virol.* 85:3842–3857.

Dutko JA, Schäfer A, Kenny AE, Cullen BR, Curcio MJ, Schafer A. 2005. Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. *Curr Biol.* 15:661–666.

Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, Hance AJ, Heidmann T, Schwartz O. 2005. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* 433:430–433.

Esnault C, Millet J, Schwartz O, Heidmann T. 2006. Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. *Nucleic Acids Res.* 34:1522–1531.

Esnault C, Priet S, Ribet D, Heidmann O, Heidmann T. 2008. Restriction by APOBEC3 proteins of endogenous retroviruses with an extracellular life cycle: ex vivo effects and in vivo “traces” on the murine IAPE and human HERV-K elements. *Retrovirology* 5:75.

Fang X, Nevo E, Han L, Levanon EY, Zhao J, Avivi A, Larkin D, Jiang X, Feranchuk S, Zhu Y, et al. 2014. Genome-wide adaptive complexes to underground stresses in blind mole rats *Spalax*. *Nat Commun.* 5:3966.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet.* 9:397–405.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, et al. 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41:D808–D815.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.* 15:1451–1455.

Guo F, Cen S, Niu M, Saadatmand J, Kleiman L. 2006. Inhibition of formula-primed reverse transcription by human APOBEC3G during human immunodeficiency virus type 1 replication. *J Virol.* 80:11710–11722.

Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH. 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell* 113:803–809.

Holmes RK, Malim MH, Bishop KN. 2007. APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci.* 32:118–128.

Horn AV, Klawitter S, Held U, Berger A, Vasudevan AAJ, Bock A, Hofmann H, Hanschmann K-MO, Trösemeyer J-H, Flory E, et al. 2014. Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. *Nucleic Acids Res.* 42:396–416.

Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, et al. 2002. The Ensembl genome database project. *Nucleic Acids Res.* 30:38–41.

- Ikeda T, Abd El, Galil KH, Tokunaga K, Maeda K, Sata T, Sakaguchi N, Heidmann T, Koito A. 2011. Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. *Nucleic Acids Res.* 39:5538–5554.
- Ikeda T, Ohsugi T, Kimura T, Matsushita S, Maeda Y, Harada S, Koito A. 2008. The antiretroviral potency of APOBEC1 deaminase from small animal species. *Nucleic Acids Res.* 36:6859–6871.
- Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N. 2002. An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22. *Genomics* 79:285–296.
- Jónsson SR, Haché G, Stenglein MD, Fahrenkrug SC, Andrésdóttir V, Harris RS. 2006. Evolutionarily conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins. *Nucleic Acids Res.* 34:5683–5694.
- Kaiser SM, Emerman M. 2006. Uracil DNA glycosylase is dispensable for human immunodeficiency virus type 1 replication and does not contribute to the antiviral effects of the cytidine deaminase APOBEC3G. *J Virol.* 80:875–882.
- Karolchik D. 2003. The UCSC Genome Browser database. *Nucleic Acids Res.* 31:51–54.
- Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, Kent WJ. 2004. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32:D493–D496.
- Kazazian HH. 2004. Mobile elements: drivers of genome evolution. *Science* 303:1626–1632.
- Koito A, Ikeda T. 2011. Intrinsic restriction activity by AID/APOBEC family of enzymes against the mobility of retroelements. *Mob Genet Elements.* 1:197–202.
- Koito A, Ikeda T. 2013. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Front Microbiol.* 4:28.
- Langlois M-A, Neuberger MS. 2008. Human APOBEC3G can restrict retroviral infection in avian cells and acts independently of both UNG and SMUG1. *J Virol.* 82:4660–4664.
- LaRue RS, Andrésdóttir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jónsson SR, Landau NR, Löchelt M, et al. 2009. Guidelines for naming nonprimate APOBEC3 genes and proteins. *J Virol.* 83:494–497.
- Lecossier D, Bouchonnet F, Clavel F, Hance AJ. 2003. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* 300:1112.
- Liddament MT, Brown WL, Schumacher AJ, Harris RS. 2004. APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo. *Curr Biol.* 14:1385–1391.
- Lindič N, Budič M, Petan T, Knisbacher BA, Levanon EY, Lovšin N. 2013. Differential inhibition of LINE1 and LINE2 retrotransposition by vertebrate AID/APOBEC proteins. *Retrovirology* 10:156.
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72:595–605.
- Luo K, Wang T, Liu B, Tian C, Xiao Z, Kappes J, Yu X-F. 2007. Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J Virol.* 81:7238–7248.
- MacDuff DA, Demorest ZL, Harris RS. 2009. AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Res.* 37:1854–1867.
- Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. 2003. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 424:99–103.
- Mariani R, Chen D, Schröfelbauer B, Navarro F, König R, Bollman B, Münk C, Nymark-McMahon H, Landau NR, Schrofelbauer B, et al. 2003. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* 114:21–31.
- Mbisa JL, Bu W, Pathak VK. 2010. APOBEC3F and APOBEC3G inhibit HIV-1 DNA integration by different mechanisms. *J Virol.* 84:5250–5259.
- Miyagi E, Opi S, Takeuchi H, Khan M, Goila-Gaur R, Kao S, Strebel K. 2007. Enzymatically active APOBEC3G is required for efficient inhibition of human immunodeficiency virus type 1. *J Virol.* 81:13346–13353.
- Muckenfuss H, Hamdorf M, Held U, Perkovic M, Löwer J, Cichutek K, Flory E, Schumann GG, Münk C. 2006. APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J Biol Chem.* 281:22161–22172.
- Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102:553–563.
- Nellåker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, Flint J, Adams DJ, Frankel WN, Ponting CP. 2012. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* 13:R45.
- Newman ENC, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM. 2005. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol.* 15:166–170.
- Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* 499:471–475.
- Prohaska KM, Bennett RP, Salter JD, Smith HC. 2014. The multifaceted roles of RNA binding in APOBEC cytidine deaminase functions. *Wiley Interdiscip Rev RNA.* 5:493–508.
- Refsland EW, Harris RS. 2013. The APOBEC3 family of retroelement restriction factors. *Curr Top Microbiol Immunol.* 371:1–27.
- Sawyer SL, Emerman M, Malik HS. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* 2:E275.
- Schumacher AJ, Haché G, Macduff DA, Brown WL, Harris RS. 2008. The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. *J Virol.* 82:2652–2660.
- Schumacher AJ, Nissley DV, Harris RS. 2005. APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast. *Proc Natl Acad Sci U S A.* 102:9854–9859.
- Severi F, Chicca A, Conticello SG. 2011. Analysis of reptilian APOBEC1 suggests that RNA editing may not be its ancestral function. *Mol Biol Evol.* 28:1125–1129.
- Smit AF. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev.* 9:657–663.
- Smit AFA, Hubley R, Green P. RepeatMasker Open-3.0, 1996–2010. Available from: <http://www.repeatmasker.org>.
- Smith HC, Bennett RP, Kizilyer A, McDougall WM, Prohaska KM. 2012. Functions and regulation of the APOBEC family of proteins. *Semin Cell Dev Biol.* 23:258–268.
- Stenglein MD, Harris RS. 2006. APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J Biol Chem.* 281:16837–16841.
- Suspène R, Aynaud M-M, Guétard D, Henry M, Eckhoff G, Marchio A, Pineau P, Dejean A, Vartanian J-P, Wain-Hobson S. 2011. Somatic hypermutation of human mitochondrial and nuclear DNA by APOBEC3 cytidine deaminases, a pathway for DNA catabolism. *Proc Natl Acad Sci U S A.* 108:4858–4863.
- Tsirigos A, Rigoutsos I. 2009. Alu and b1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. *PLoS Comput Biol.* 5:e1000610.
- Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D. 2007. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci U S A.* 104:18613–18618.
- Zarank AW, Levanon EY, Zecharia T, Clegg T, Church GM. 2010. A survey of genomic traces reveals a common sequencing error, RNA editing, and DNA editing. *PLoS Genet.* 6:e1000954.
- Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320.
- Zhang J, Webb DM. 2004. Rapid evolution of primate antiviral enzyme APOBEC3G. *Hum Mol Genet.* 13:1785–1791.