

Research article

The artiodactyl *APOBEC3* innate immune repertoire shows evidence for a multi-functional domain organization that existed in the ancestor of placental mammals

Rebecca S LaRue¹, Stefán R Jónsson^{1,2}, Kevin AT Silverstein³, Mathieu Lajoie⁴, Denis Bertrand⁴, Nadia El-Mabrouk⁴, Isidro Hötzel⁵, Valgerdur Andrésdóttir², Timothy PL Smith⁶ and Reuben S Harris*¹

Address: ¹Department of Biochemistry, Molecular Biology and Biophysics, Institute for Molecular Virology, Beckman Center for Genome Engineering, University of Minnesota, Minneapolis, Minnesota 55455, USA, ²University of Iceland, Institute for Experimental Pathology, Keldur v/Vesturlandsveg, 112 Reykjavík, Iceland, ³Masonic Cancer Center, Biostatistics and Bioinformatics Group, University of Minnesota, Minneapolis, Minnesota 55455, USA, ⁴DIRO, Université de Montréal, Montréal, Québec, H3C 3J7, Canada, ⁵Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, Washington 99164-7040, USA and ⁶USDA/ARS US Meat Animal Research Center, Genetics and Breeding Research Unit, PO Box 166, Clay Center, Nebraska 68933-0166, USA

Email: Rebecca S LaRue - larue005@umn.edu; Stefán R Jónsson - stefanjo@hi.is; Kevin AT Silverstein - silve023@umn.edu; Mathieu Lajoie - mathieu.lajoie@gmail.com; Denis Bertrand - dbertran78@gmail.com; Nadia El-Mabrouk - mabrouk@iro.umontreal.ca; Isidro Hötzel - ihotzel@mac.com; Valgerdur Andrésdóttir - valand@hi.is; Timothy PL Smith - Tim.Smith@ars.usda.gov; Reuben S Harris* - rsh@umn.edu

* Corresponding author

Published: 18 November 2008

Received: 13 August 2008

BMC Molecular Biology 2008, **9**:104 doi:10.1186/1471-2199-9-104

Accepted: 18 November 2008

This article is available from: <http://www.biomedcentral.com/1471-2199/9/104>

© 2008 LaRue et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: APOBEC3 (A3) proteins deaminate DNA cytosines and block the replication of retroviruses and retrotransposons. Each A3 gene encodes a protein with one or two conserved zinc-coordinating motifs (Z1, Z2 or Z3). The presence of one A3 gene in mice (Z2–Z3) and seven in humans, A3A–H (Z1a, Z2a–Z1b, Z2b, Z2c–Z2d, Z2e–Z2f, Z2g–Z1c, Z3), suggests extraordinary evolutionary flexibility. To gain insights into the mechanism and timing of A3 gene expansion and into the functional modularity of these genes, we analyzed the genomic sequences, expressed cDNAs and activities of the full A3 repertoire of three artiodactyl lineages: sheep, cattle and pigs.

Results: Sheep and cattle have three A3 genes, A3Z1, A3Z2 and A3Z3, whereas pigs only have two, A3Z2 and A3Z3. A comparison between domestic and wild pigs indicated that A3Z1 was deleted in the pig lineage. In all three species, read-through transcription and alternative splicing also produced a catalytically active double domain A3Z2–Z3 protein that had a distinct cytoplasmic localization. Thus, the three A3 genes of sheep and cattle encode four conserved and active proteins. These data, together with phylogenetic analyses, indicated that a similar, functionally modular A3 repertoire existed in the common ancestor of artiodactyls and primates (i.e., the ancestor of placental mammals). This mammalian ancestor therefore possessed the minimal A3 gene set, Z1–Z2–Z3, required to evolve through a remarkable series of eight recombination events into the present day eleven Z domain human repertoire.

Conclusion: The dynamic recombination-filled history of the mammalian A3 genes is consistent with the modular nature of the locus and a model in which most of these events (especially the expansions) were selected by ancient pathogenic retrovirus infections.

Background

Mammalian APOBEC3 (A3) proteins have the capacity to potently inhibit the replication of a diverse set of reverse-transcribing mobile genetic elements [1-5]. Susceptible exogenous retroelements include lentiviruses (HIV-1, HIV-2, several strains of SIV and FIV), alpharetroviruses (RSV), betaretroviruses (MPMV), gammaretroviruses (MLV), deltaretroviruses (HTLV), foamy viruses and the hepadnavirus HBV (*e.g.*, [6-14]). Susceptible endogenous retroelements include the yeast retrotransposons Ty1 and Ty2, the murine endogenous retroviruses MusD and Pmv, the murine intracisternal A particle (IAP), the porcine endogenous retrovirus PERV and, potentially, extinct elements such as chimpanzee PtERV1 and human HERV-K, all of which require long-terminal repeats (LTRs) for replication [15-23]. In addition, some A3 proteins can also inhibit L1 and its obligate parasite Alu, retrotransposons that replicate by integration-primed reverse transcription [24-30]. An overall theme is emerging in which most – if not all – retroelements can be inhibited by at least one A3 protein.

However, it is now equally clear that the retroelements of any given species have evolved mechanisms to evade restriction by their host's A3 protein(s). For instance, HIV and SIV use Vif to trigger a ubiquitin-dependent degradation mechanism, foamy viruses use a protein called Bet for an imprecisely defined inhibitory mechanism and some viruses such as MPMV, HTLV and MLV appear to employ a simple avoidance mechanism (*e.g.*, [6,31-34]). Thus, it appears that all 'successful' retroelements have evolved strategies to resist restriction by the A3 proteins of their hosts.

The defining feature of the A3 family of proteins is a conserved zinc(Z)-coordinating DNA cytosine deaminase motif, H-x₁-E-x₂₅₋₃₁-C-x₂₋₄-C (x indicates a non-conserved position [35,36]). The A3 Z domains can be grouped into one of three distinct phylogenetic clusters – Z1, Z2 or Z3. (Figure 1 & Additional File 1). The Z-based classification system, proposed originally by Conticello and coworkers [35], was revised recently through a collaborative effort [37]. From hereon, the new A3 nomenclature system will be used. Z1 and Z2 proteins have a SW-S/T-C-x₂₋₄-C motif, whereas Z3 proteins have a TW-S-C-x₂-C motif. Z1 and Z2 proteins can be further distinguished by H-x₁-E-x₅-X-V/I and H-x₁-E-x₅-W-F motifs, respectively. Z1 proteins also have a unique isoleucine within a conserved RLY motif located C-terminal to the zinc-coordinating residues. At least one protein of each of the Z classes and nearly all identified A3 proteins have exhibited single-strand DNA cytosine deaminase activity. For instance, human A3F, A3G and A3H possess catalytically competent Z2, Z1 and Z3 domains, respectively (*e.g.*, [38-41]).

We previously reported a double-domain A3Z2-Z3 gene (formerly called A3F) from the artiodactyls, sheep (*Ovis aries*), cattle (*Bos taurus*) and pigs (*Sus scrofa*) [42]. However, the fact that mammals have varying numbers of A3 genes (*e.g.*, 7 in humans and only 1 in mice) led us to wonder whether additional A3 genes would be present in artiodactyls. To address this point and to learn more about the evolution and functionality of A3 genes in mammals, we sequenced and characterized the full A3 repertoire of sheep and pigs. Here, we demonstrated that sheep and cattle actually have three A3 genes, A3Z1, A3Z2 and A3Z3, with a conserved potential to encode at least four active and distinct proteins (A3Z1, A3Z2, A3Z3 and A3Z2-Z3). We further showed that porcine lineage has a deletion of the orthologous A3Z1 gene and the capacity to encode only three proteins. These data enabled us to deduce that the common ancestor of artiodactyls and primates possessed an A3 repertoire consisting of three Z domains (Z1, Z2 and Z3). Our data further suggested an evolutionary model in which most of the human A3 gene expansion occurred more than 25 million years ago, during early primate evolution and possibly even associated with pathogen-induced population bottlenecks.

Results

Sheep and cattle have three A3 genes with a Z1-Z2-Z3 organization

We previously used degenerate PCR, RACE and database mining to identify a cDNA for sheep A3Z2-Z3 (formerly called A3F; [42]). However, because humans have seven A3 genes and mice have only one, we postulated that artiodactyls such as sheep and cattle might have an intermediate number. To address this possibility unambiguously, we sequenced the entire sheep A3 genomic locus. First, a sheep A3Z2-Z3 cDNA was hybridized to a sheep BAC library to identify corresponding genomic sequence. Second, hybridization-positive BACS were screened by PCR for those that also contain the conserved flanking genes *CBX6* and *CBX7*. One BAC was identified that spanned the entire *CBX6* to *CBX7* region, and it was sheared, sub-cloned, shotgun sequenced, assembled and analyzed (Methods).

DNA sequence analyses revealed that the sheep genomic locus contained another A3 gene between *CBX6* and A3Z2 (Figure 2A). This gene was called A3Z1, because it had sequence characteristics of a Z1-type A3 protein. We therefore concluded that sheep have three A3 genes and, importantly, that each mammalian A3 Z-type was present. This conclusion was supported by the bovine genome assembly, which was released during the course of our studies and showed that cattle also have a sheep-like, three gene A3 repertoire (Figure 2A; Btau_4.0 <http://www.hgsc.bcm.tmc.edu/projects/bovine/>). The predicted A3Z1 coding sequences of sheep and cattle are 86% iden-

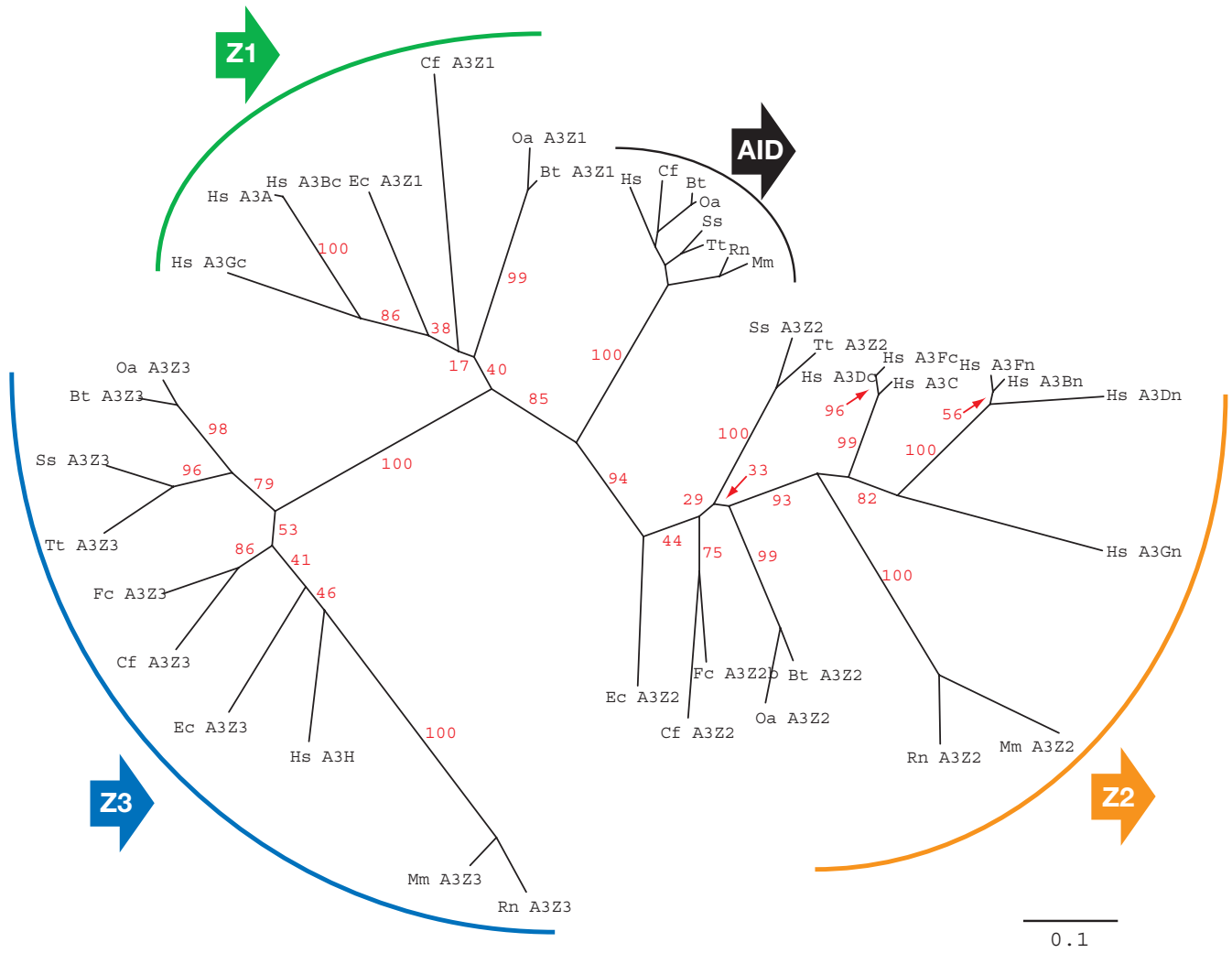


Figure 1
The mammalian A3 Z domains form three distinct phylogenetic groups. Bootstrap values are indicated in red. The scale bar represents 0.1 nucleotide changes per codon. See the Methods for details. Abbreviations for mammals: Hs = human, Bt = cow, Oa = sheep, Ss = pig, Tt = peccary, Ec = horse, Cf = dog, Fc = cat, Mm = mouse and Rn = rat. Other abbreviations: n = amino terminal domain and c = carboxy terminal domain.

tical, consistent with the fact that these two ruminant artiodactyls shared a common ancestor approximately 14–25 million years ago (MYA) [43-45].

The pig has two A3 genes with a Z2–Z3 organization

PCR reactions failed to identify an A3Z1-like gene in pigs. Since pigs and cattle/sheep last shared a common ancestor approximately 70–80 MYA [43,44], we considered the possibility that the negative PCR result was not a technical failure and that pigs might actually have a different A3 repertoire. Again, to unambiguously address this possibility, the pig A3 genomic locus was sequenced in entirety. A porcine BAC library was probed with pig A3Z2-Z3 cDNA

and two hybridization-positive BACS were shotgun sequenced. The sequence assemblies revealed that pigs have only two A3 genes A3Z2 and A3Z3 between CBX6 and CBX7 (Figure 2A).

The cattle, sheep and pig A3 locus genomic sequences were compared using dotplot analyses (Figure 2B & Additional File 2). A 22 kb discontinuity was detected between the cow and the pig sequences. The sheep and pig genomic sequences aligned similarly. Multiple (likely inactive) retroelements were found to flank A3Z1 in sheep and cattle. Two were particularly close to the ends of the 22 kb A3Z1 region, a LINE/L1 and a SINE/tRNA-Glu. It is

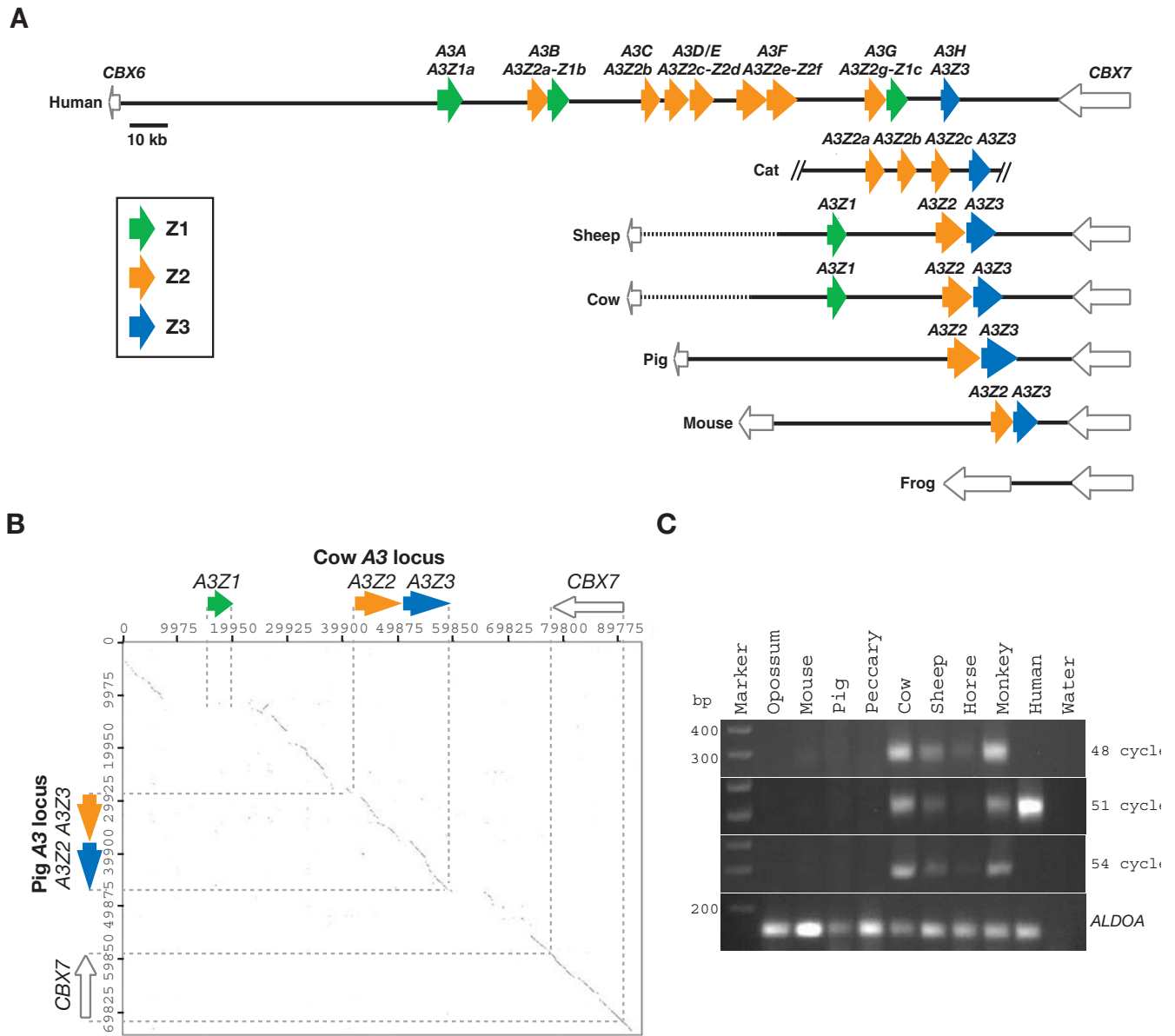


Figure 2
The A3 genomic repertoire of sheep, cattle and pigs. (A) An illustration of the A3 genes of the indicated mammals. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The conserved flanking genes CBX6 and CBX7 are shown and the scale is indicated. Solid lines represent finished sequence and dotted lines represent gaps or incomplete regions. Non-mammalian vertebrates such as frogs lack A3 genes. (B) A dotplot analysis shows A3Z1 in cattle but not in pig genomic sequence. The x- and y-axis numbers designate nucleotide positions within the indicated genomic consensus sequences. (C) PCR analysis of genomic DNA from the indicated species showing that a 250–256 bp Z1-specific amplicon can only be obtained from a subset of mammals. The human 51 and 54 cycle amplicons were too abundant to be run on the same gel. Monkey genomic DNA is from the African green monkey. The ALDOA gene was used as a positive control (115 bp).

possible that one of these elements mediated a simple direct repeat recombination event that deleted the A3Z1 region in pigs. However, we were unable to identify such a causative retroelement in the pig genomic sequence.

To begin to address whether the potential A3Z1 deletion in pigs occurred recently (e.g., a rare deletion fixed by selective breeding) or whether it was more ancient, we asked whether a non-domesticated, distant relative of the

pig, the collared peccary (*Tayassu tajacu*), has an *A3Z1* gene. Lineages leading to present-day domesticated pigs and the peccary diverged approximately 25–35 MYA [43]. A pan-species, *A3Z1* PCR primer set was developed and used in these experiments. In contrast to human, African green monkey, horse, cow and sheep genomic DNA which yielded a 250–256 bp *Z1*-specific PCR products confirmable by DNA sequencing, the genomic DNA of domesticated pig, the collared peccary, mice and opossum failed to yield a product even after 54 cycles (Figure 2C). A highly conserved gene, *ALDOA*, was used as a PCR control to demonstrate the integrity of the genomic DNA samples.

Interestingly, *Z1* PCR product sequencing and recently released EST sequences revealed that the related hoofed mammal, the horse, also has a *Z1*-type *A3* gene (Figure 1C & Additional File 3). Two-'toed' hoofed animals such as sheep, cattle and pigs belong to the ungulate order artiodactyla (even-toe number), and one-'toed' hoofed animals such as horses belong to the ungulate order perissodactyla (odd-toe number). Since these two ungulate orders diverged approximately 80–90 MYA [43,44] and both have species with *Z1*-type *A3* genes, it is highly likely that the common ancestor also had an *A3Z1* gene (as well as *A3Z2* and *A3Z3* genes). It is therefore highly unlikely that an *A3Z1* gene independently appeared at the same genomic position in artiodactyls, perissodactyls and primates. Rather, all of the data support a model where a common ancestor of the domesticated pig and the collared peccary experienced a 22 kb deletion that resulted in the loss of *A3Z1* (*i.e.*, a divergent evolutionary model). Furthermore, since artiodactyls, perissodactyls and humans shared a common ancestor approximately 80–120 MYA [43,44], the presence of *Z1*-type *A3* genes in both the primate and the artiodactyl limbs of the mammalian tree is also most easily explained by common ancestry. Thus, our combined datasets indicated that this ancestor possessed a full *A3 Z* repertoire, with one of each type of *Z* domain (*Z1*, *Z2* and *Z3*), the minimal substrate required to evolve into the present-day eleven *Z* domain human *A3* locus (discussed further below).

The artiodactyl *A3Z2* and *A3Z3* genes combine to encode 3 distinct mRNAs and proteins

We previously characterized several activities of the double-domain *A3Z2-Z3* protein from cattle, sheep and pigs [42]. While re-confirming the 5' and 3' ends of the *A3Z2-Z3* transcripts by RACE, we discovered two interesting variants that were conserved between these three species. First, using sheep and cattle PBMC or cell line cDNA (FLK and MDBK, respectively), 3' RACE frequently produced a smaller than expected fragment. The sequence of this fragment indicated the existence of a short 1037 bp transcript due to premature termination 329 or 330 nucleotides into intron 4 for sheep and cattle, respectively (Figure 3). This

truncated transcript was readily amplified from sheep and cattle PBMCs and represented by existing EST sequences (Additional File 3 and data not shown). Therefore, this novel transcript was predicted to result in a single-domain *Z2* protein, *A3Z2*, with a length of 189 and 202 amino acids for sheep and cattle, respectively (Figure 3 & Additional File 3). A pig *A3Z2* transcript was also identified by RACE and EST sequences but, in contrast to sheep and cattle, exon 4 was spliced to two additional exons before terminating prematurely (Figure 3 & Additional File 3). As a consequence, pig *A3Z2* was predicted to be 265 amino acids. These analyses indicated that artiodactyls have the capacity to express a single domain *A3Z2* protein, in contrast to what we had deduced previously [42].

Second, 5' RACE data and cattle and pig EST sequences suggested that yet another mechanism served to broaden the coding potential of the artiodactyl *A3* locus (Additional File 3 and data not shown). Several transcripts appeared to originate from the region immediately upstream of *A3Z3*, whereas our prior studies had only detected transcripts originating upstream of *A3Z2* [42]. A comparison of cDNA and genomic sequences revealed the presence of an exon in this location (*A3Z3* exon 1 in Figure 3). Transcripts initiating here produced 941 (sheep), 964 (cow) or 1003 (pig) nucleotide messages. The resulting *A3Z3* protein was predicted to be 206 residues for sheep and cattle and 207 for pigs (Figure 3).

The *A3Z3* mRNA data strongly suggested the existence of an internal promoter. This was supported by cis-regulatory element prediction algorithms, which identified a conserved interferon-stimulated response element (ISRE) upstream of *A3Z3*, as well as upstream of *A3Z2* (Figure 3 & Additional File 4). These ISREs were strikingly similar to those located in the promoter regions of human *A3DE*, *A3F* and *A3G*, supporting the likelihood that interferon-inducibility is a conserved feature of many mammalian *A3* genes (*e.g.*, [8,46-49]). These putative ISREs significant similarity to functional elements in known interferon-inducible genes *ISG54* and *ISG15* [50-52]. We also predicted binding sites for another well-known transcription factor, Sp1, upstream of the *A3Z3* transcription start site. This activator was also recently reported for human and cat *A3* genes ([53,54]; LaRue & Harris, data not shown).

Together with our previous data on the double domain *A3* protein of these artiodactyl species, *A3Z2-Z3* [42], these expression and promoter data revealed that two single-domain *A3* genes can readily encode at least three distinct proteins – *A3Z2*, *A3Z3* and *A3Z2-Z3*. A similar strategy may also be used by rodents, which also have an *A3* gene with *Z2* and *Z3* domains. A similar modularity was reported recently for the cat *A3* locus, where two single domain *A3* genes combined to produce a functional dou-

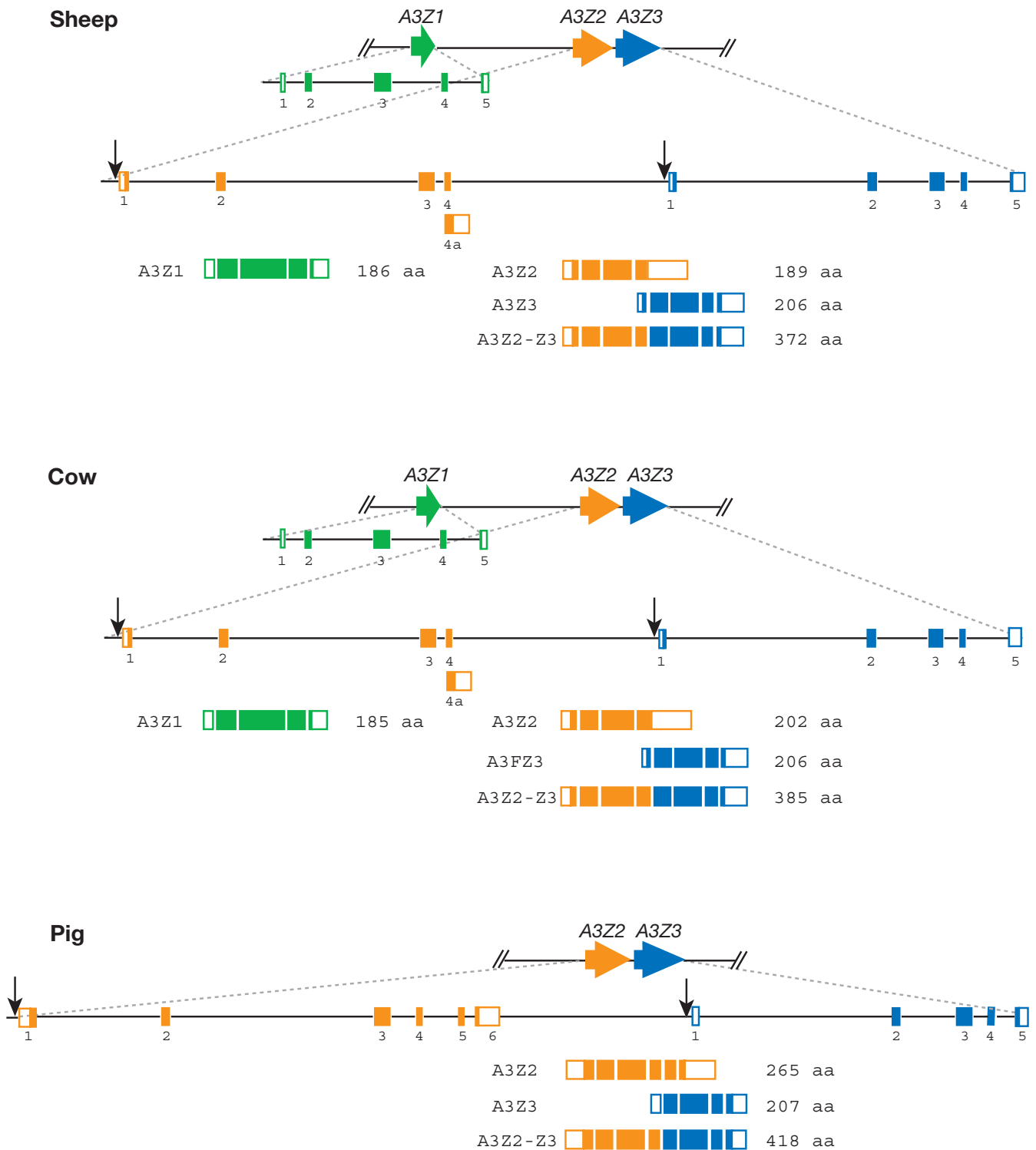


Figure 3
The coding potential of the sheep, cow and pig A3 genes. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The exons are shown below the gene schematics with coding regions represented by filled boxes and untranslated regions by open boxes. The gene schematics and exon blow-ups are drawn to scale. Arrows indicate approximate positions of predicted ISREs (Additional File 4). See the main text and the Methods for details.

ble-domain A3 protein [54]. We suggest that combining single-domain A3s to yield functionally unique double-domain proteins may be a general strategy used by many mammals to bolster their A3-dependent innate immune defenses.

All four artiodactyl A3 proteins – A3Z1, A3Z2, A3Z3 and A3Z2-Z3 – elicit DNA cytosine deaminase activity

All currently described A3 proteins have elicited single-strand DNA cytosine to uracil deaminase activity in one or more assays (e.g., [24,41,42,54-59]). For instance, we showed that the artiodactyl A3Z2-Z3 proteins could catalyze the deamination of *E. coli* DNA and retroviral cDNA [42]. However, catalytic mutants indicated that only the N-terminal Z2 domain of cow, sheep and pig A3Z2-Z3 was active. This observation contrasted with data for the double-domain human A3B, A3F and A3G proteins, where the C-terminal domain clearly contains the dominant active site (e.g., [30,38-40,42,60]). Nevertheless, these datasets suggested that the double-domain A3 proteins have separated function, with one domain predominantly serving as a catalytic center and the other as a regulatory center.

However, a recent study with human A3B indicated that both Z domains have the potential to be catalytically active [61]. It was therefore reasonable to ask whether the single domain A3Z2 and A3Z3 proteins of artiodactyls would be capable of DNA cytosine deamination in an *E. coli*-based activity assay. Elevated frequencies of rifampicin-resistance (Rif^R) mutations in *E. coli* provide a

quantitative measure of the intrinsic A3 protein DNA cytosine deaminase activity (e.g., [38,40,56,57]). In contrast to full-length cow A3Z2-Z3, which triggered a modest 2-fold increase in the median Rif^R mutation frequency over the vector control, non-induced levels of cow A3Z2 caused a large 50-fold increase (Figure 4A). The pTrc99-based vector used in these studies has an IPTG-inducible promoter, and induced levels of cow A3Z2 prevented *E. coli* growth, presumably through catastrophic levels of DNA cytosine deamination. In contrast, induced levels of sheep or pig A3Z2 proteins were not lethal, but their expression also caused significant increases in the median Rif^R mutation frequency (Additional File 5 and LaRue & Harris, data not shown). Thus, as anticipated by our prior studies, the A3Z2 proteins of cattle, sheep and pigs showed intrinsic DNA cytosine deaminase activity.

We were therefore surprised that induced levels of the cow single-domain protein A3Z3 also caused a significant 4-fold increase in the median Rif^R mutation frequency (Figure 4B). This result contrasted with the related Z3 protein of humans, A3H, which appeared inactive in this assay (Figure 4B & Additional File 3). However, it is worth noting that other Z3-type A3 proteins, a different human A3H variant, African green monkey A3H, rhesus macaque A3H and cat A3Z3 (formally A3H), all showed evidence for DNA deaminase activity in the *E. coli*-based mutation assay and/or in retrovirus infectivity assays [41,54,62,63]. Thus, our intended human A3H control appears to be the exception rather than the rule and that the single-domain

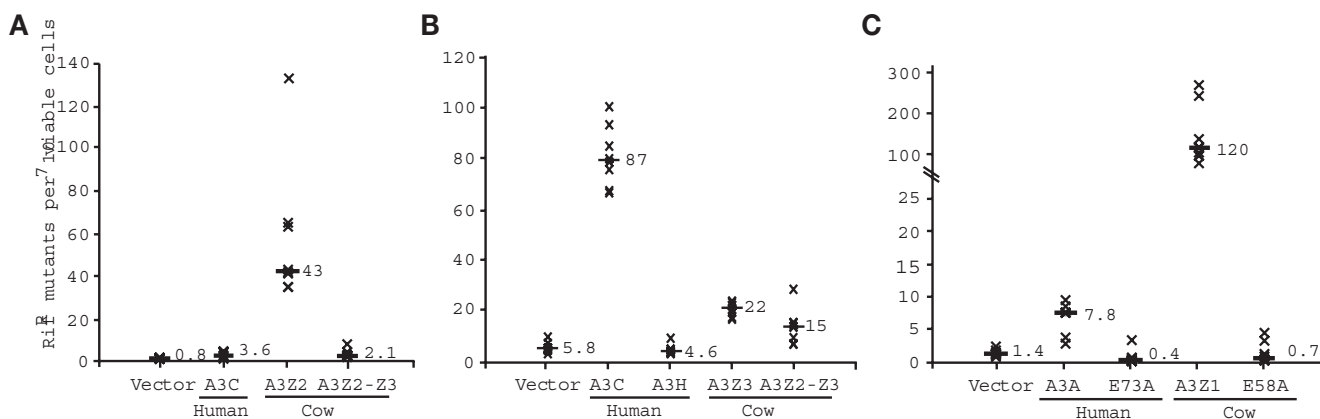


Figure 4
The artiodactyl A3 proteins catalyze DNA cytosine deamination. (A) Cow A3Z2 triggers a strong mutator phenotype in *E. coli*. Rif^R mutation frequency of 4–6 independent bacteria cultures expressing basal levels of indicated A3 proteins (each X represents data from a single culture). To facilitate comparisons, the median mutation frequency is indicated for each condition. (B) Cow A3Z3 triggers a modest mutator phenotype in *E. coli*. Labels and conditions are similar to those in panel (A), except IPTG was used to induce protein expression. (C) Non-induced cow A3Z1 triggers a strong mutator phenotype in *E. coli*, which is completely abrogated by substituting the catalytic glutamate (E58) for alanine. Labels are similar to those in panel (A).

A3Z3 protein of artiodactyls is capable of DNA cytosine deaminase activity.

We also observed that the artiodactyl A3Z1 protein was capable of robust DNA cytosine deaminase activity (*e.g.*, Figure 4C and LaRue and Harris, unpublished data). This result was fully anticipated based on the fact that the related Z1 domain proteins of humans A3A, A3B and A3C are catalytically active [24,30,61,64]. However, it is worth noting three observations suggesting that cow A3Z1 is the most active of all reported A3 proteins. First, we were never able to directionally clone (even non-induced) A3Z1 of sheep or cattle into pTrc99A, which has a leaky promoter. Second, we were only able to topoisomerase-clone cow A3Z1 in a direction opposite to the *lac* promoter ($n > 12$). Finally, even with cow A3Z1 in the promoter-opposing orientation in the topoisomerase cloning plasmid, we observed 100-fold increases in Rif^R mutation frequency in the *E. coli*-based mutation assay that were fully dependent on the catalytic glutamate E58A (presumably due to expression from a cryptic promoter; Figure 4C). To summarize this section, all four of the A3 proteins of artiodactyls demonstrated intrinsic DNA cytosine deaminase activity.

A3Z1, A3Z2, A3Z3 and A3Z2-Z3 differentially localize in cells

Fluorescent microscopy was used to examine the subcellular distribution of each of the artiodactyl A3 proteins fused to GFP. Like the human A3 proteins, which each have unique overall subcellular distributions, we imagined that distinct localization patterns might correlate with differential functions. For instance, the first column of Figure 5 shows representative images of live HeLa cells expressing human A3F-GFP, A3A-GFP, A3C-GFP and A3H-GFP, which predominantly localize to the cytoplasm, cell-wide with a nuclear bias, cell-wide and cell-wide with a clear nucleolar preference, respectively. Cow A3Z1-GFP showed an indiscriminate cell-wide distribution similar to that of human A3A-GFP and GFP alone (Figure 5, second row and data not shown).

As shown previously, cattle and pig A3Z2-Z3-GFP localize to the cytoplasm, with some cells showing bright aggregates (Figure 5, row 1; [19,42]). Cattle and pig A3Z2 also appeared predominantly cytoplasmic, but a significant fraction clearly penetrated the nuclear compartment (row 3). The subcellular distribution of cattle and pig A3Z2 differed from the similarly sized Z2 protein human A3C, which was cell-wide, and it is therefore likely that an active process underlies the cytoplasmic bias of the artiodactyl A3Z2 proteins. Interestingly, the A3Z3 proteins of cattle and sheep, like human A3H, localized cell-wide with clear accumulations in the nucleoli (row 4). Similar data were obtained using these GFP fusion constructs in live cattle MDBK cells and in live pig PK15 cells (LaRue &

Harris, data not shown). These fluorescent microscopy observations demonstrated that all of the artiodactyl A3 proteins can be expressed in mammalian cells and that they have both distinct and overlapping subcellular distributions.

The artiodactyl A3 genes show evidence for positive selection

Many human, non-human primate and feline A3 genes show signs of strong positive selection, which can be interpreted as evidence for a history filled with pathogen conflicts [41,54,65,66]. However, given the relative stability of the artiodactyl A3 locus, at least in terms of gene number, we wondered whether the artiodactyl A3 genes might be under less intense selective pressure (perhaps even neutral or negative). This possibility was assessed using two methods to compare the number of mutations that resulted in amino acid replacements to the number that were silent between pairs of artiodactyl species. This ratio of replacement (dN) to silent (dS) mutations yields an omega (ω) value, which if greater than one is indicative of positive selection, if equal to one of neutral selection and if less than one of negative selection. We focused these analyses on the single exon that encodes the conserved Z domain to minimize potentially confounding effects from recombination.

We first generated a combined phylogeny for each distinct A3 Z domain and its inferred ancestral sequences (Additional File 6). Using the PAML free ratio model, the artiodactyl A3Z1 and the A3Z2 genes appeared to be under a weak negative selection pressure, with ω values uniformly below one (Additional File 6). Similarly, since the existence of the last common ancestor of cattle and sheep or of the pig and peccary, the artiodactyl A3Z3 genes showed evidence for weak negative selection pressure (Additional File 6C). However, a comparison of the inferred ancestral ruminant sequence with the inferred porcine sequence yielded a ω value of 1.5, suggesting that the ancestor(s) of modern day artiodactyls may have experienced intermittent positive selection (Additional File 6C). These values were not as high as those for primate A3Z3 (A3H data originally reported by [41] and re-calculated here with a representative clade shown in Additional File 6C). Moreover, all of these data contrasted sharply with the artiodactyl and primate AID genes, which are under an obvious strong negative selection pressure presumably for essential functions in antibody diversification.

However, because the free ratio model averages all possible sites and has a tendency to underestimate instances of positive selection, we subsequently used PAML NsSites to do a more focussed examination of artiodactyl A3 Z domain variation. Several distinct selection models were used (M2 and M8 and two codon frequency models F61 and F3 \times 4), and each yielded significant signs of positive

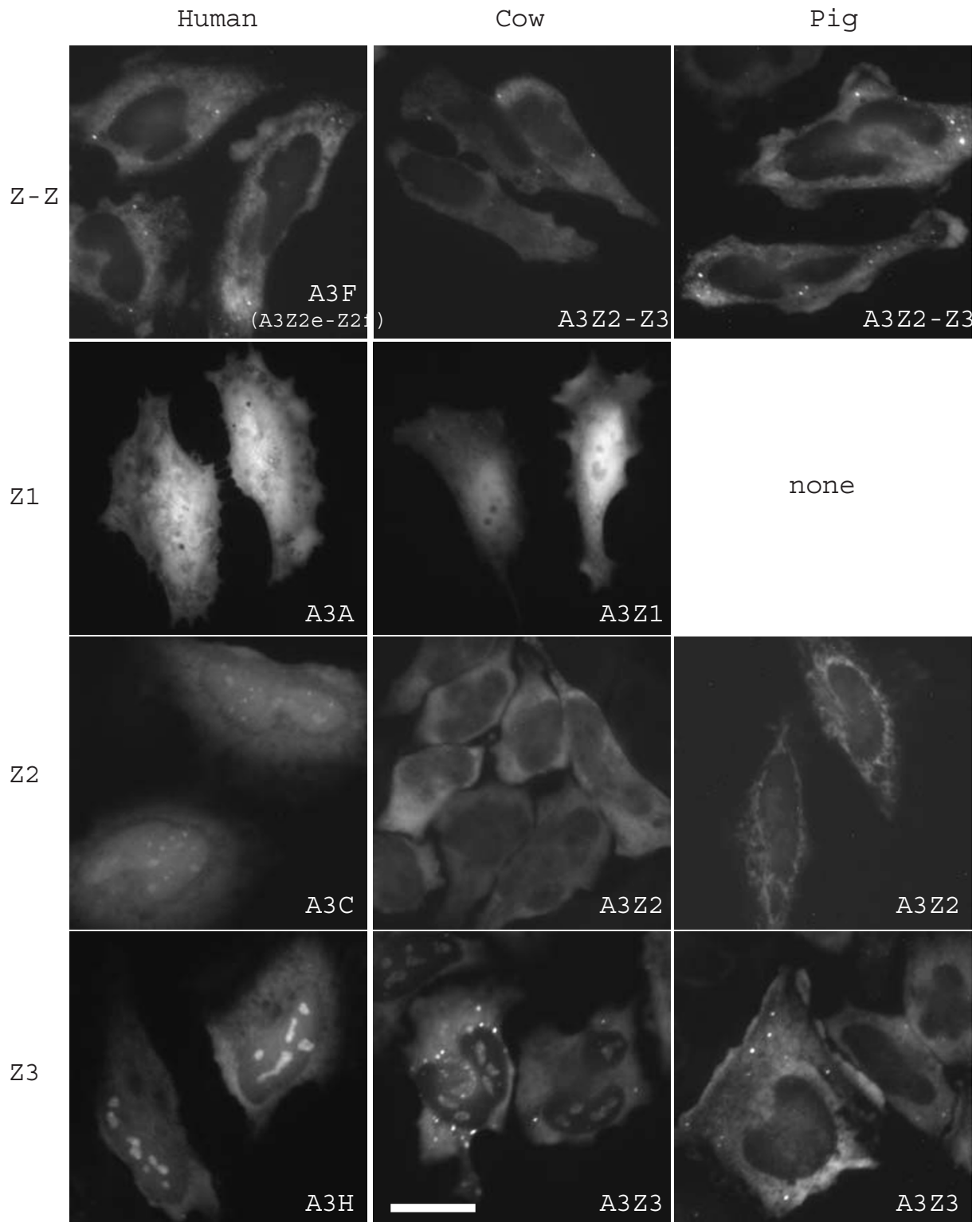


Figure 5
The subcellular distribution of cow and pig A3 proteins in comparison to human A3 proteins with similar Z domains. Representative images of live HeLa cells expressing the indicated A3-GFP fusion proteins are shown. The scale bar represents 25 μm.

selection (Table 1; see Methods for procedural details and Additional File 3 for sequence information). The Z3 domain A3 genes of sheep, cattle, pig, peccary and horse showed the highest dN/dS ratios, ranging from 4.4 to 5.8 and indicating that 22–31% of the residues were subjected to positive selection. Lower but still significant positive dN/dS ratios were obtained for the Z2 domain A3 genes (1.7 to 2.3 with 33 to 46% of the residues under positive selection). Moreover, together with available dog and horse Z1 sequences, the Z1 A3 genes of cattle and sheep showed intermediate degrees of positive selection, with dN/dS ratios of 2.5 to 3.9 and 28 to 50% of the residues under some degree of positive selection (Table 1 & Additional File 3). Thus, similar to most other mammals analyzed to date, the artiodactyl A3 genes have been subjected to strong evolutionary pressure (see Discussion).

A3 Z domain distribution in mammals

Our studies strongly indicated that the present-day A3 locus of sheep and cattle resembles one that existed in the common ancestor of placental mammals, consisting of precisely one of each of the three phylogenetically distinct Z domains: Z1, Z2 and Z3 (Figure 6; also see Figure 1 & Additional File 3). Molecular phylogenetic data helped us infer that such a common ancestor existed approximately 100–115 MYA [43,44]. However, the bulk of the primate A3 gene expansion most likely occurred more recently because the main branches leading to rodents and humans split 90–110 MYA. It is therefore likely that rodents lost a Z1 A3 gene after branching off of the main mammalian tree (like pigs, cats and some humans; see Figure 6 & Discussion). Moreover, the recently published draft of the rhesus macaque genome helped to further whittle-down when the bulk of the primate-specific

expansion occurred, because these animals also possess a human-like A3 gene repertoire (Figure 6; [41,67,68] and our unpublished data). Thus, since the human and macaque lineages diverged approximately 25 MYA [43,67,69], the massive expansion from the inferred sheep/cow-like Z1-Z2-Z3 A3 gene set to a locus resembling the present-day human repertoire must have occurred within a relatively short 65–85 million year period (indicated by an asterisk in Figure 6).

A minimum of 8 recombination events were required to generate the present-day human A3 locus from the common ancestor of artiodactyls and primates

The inferred ancestral Z1-Z2-Z3 locus was used as a starting point to deduce the most likely evolutionary scenario that transformed it into the much larger eleven Z domain human A3 repertoire. Two types of recombination events were considered, tandem duplications (obviously required for A3 gene expansion) and deletions. Self-similarities in the DNA sequence of the human A3 locus provided strong evidence for prior tandem duplications by unequal crossing-over (for more details on tandem duplication modeling see [70,71]). This mode of evolution is also supported by the fact that the human A3 locus contains many retroelements that could serve as substrates for homologous recombination [35]. Since our present studies showed that the Z domains are highly modular and capable of individual function, they were considered as the core units for duplications in our inference procedures (i.e., an unequal cross-over event can simultaneously duplicate one or more tandemly arranged Z-domains and associated flanking sequences). Similarly, deletions could involve one or more Z domains and result from unequal crossing-over or intra-chromosomal events.

Table 1: Evidence for positive selection in the artiodactyl Z domains.

Z domain ^a	Codon frequency model ^b	Comparison of null and positive selection models ^c	Significance	Tree length ^d	dN/dS (%) ^e
Z1	F6I	M1–M2	p = 0.01	2.6	2.5 (50)
		M7–M8	p = 0.01	2.6	2.5 (50)
	F3 × 4	M1–M2	p = 0.04	3.9	3.9 (28)
		M7–M8	p = 0.02	3.9	3.9 (33)
Z2	F6I	M1–M2	p = 0.005	2.4	2.3 (46)
		M7–M8	p = 0.004	2.4	2.3 (45)
	F3 × 4	M1–M2	p = 0.3	3.0	1.7 (27)
		M7–M8	p = 0.04	3.0	1.7 (33)
Z3	F6I	M1–M2	p < 0.001	2.4	4.5 (30)
		M7–M8	p < 0.001	2.4	4.4 (31)
	F3 × 4	M1–M2	p < 0.001	3.1	5.8 (22)
		M7–M8	p < 0.001	3.1	5.7 (23)

^aOnly the sequences of Z domain-encoding exons were used in these analyses (see Table S1 for GenBank accessions).

^bTwo different codon frequency models were used to minimize potentially artificial results.

^cLikelihood ratio tests were done to compare the null models M1 and M7 to each positive selection model M2 and M8, respectively, using PAML NsSites.

^dTree length provides a measure of nucleotide substitutions per codon along all combined phylogenetic branches.

^eThe percentage of all codons influenced by positive selection is indicated in parentheses.

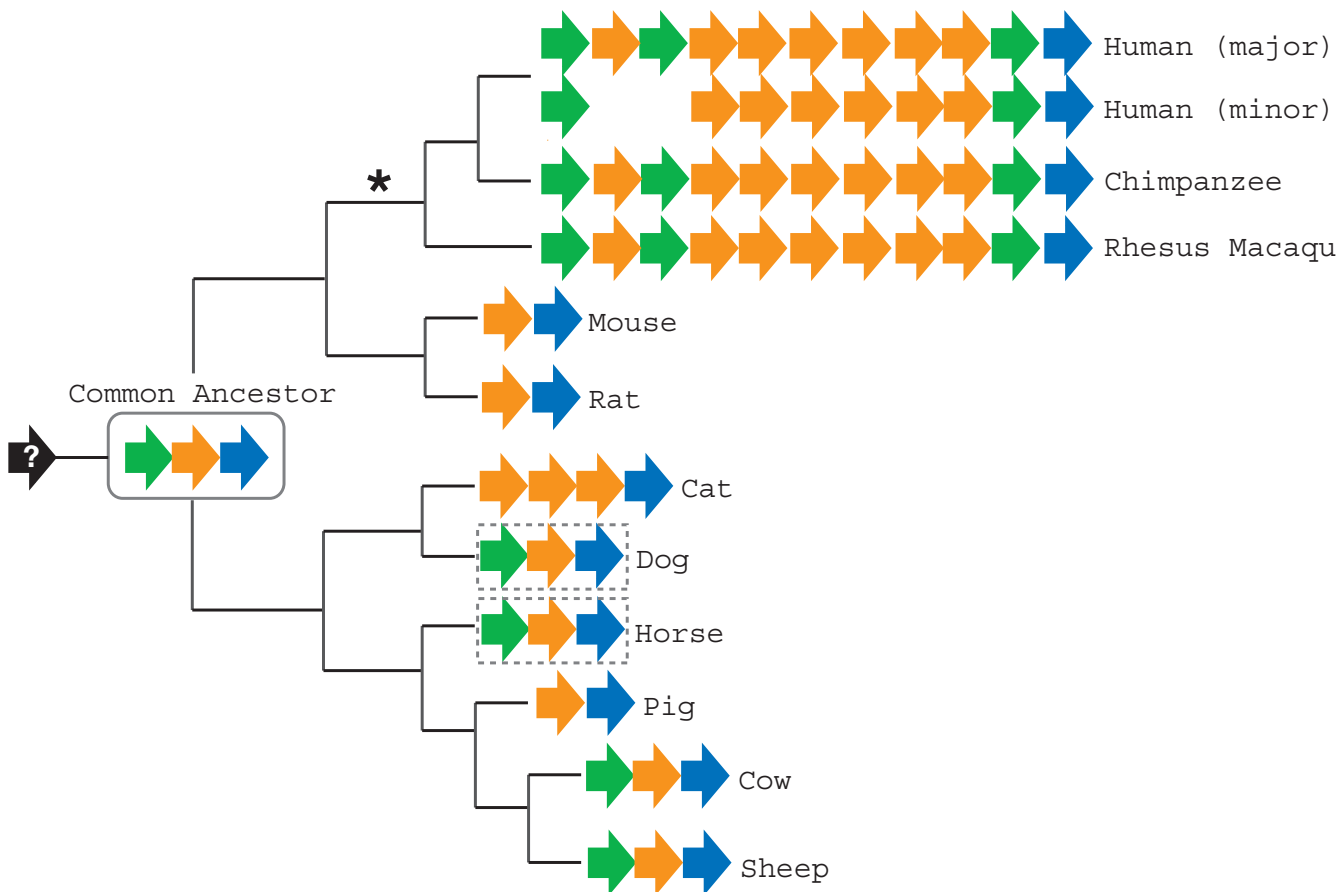


Figure 6
The distribution of A3 Z domains in mammals. The common ancestor of the indicated placental mammals was inferred to have a Z1-Z2-Z3 A3 gene repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. A question mark specifies the original AID-like ancestor. An asterisk indicates the likely period in which the bulk of the primate A3 gene expansion occurred (see main text, Figure 7 and Additional File 7). Some humans are A3B deficient (minor allele; [75]). The boxed A3 Z domain repertoires constitute the minimal set inferred from incomplete genomic sequences and EST data (Additional File 3).

An 8-event model for human A3 Z domain history is shown in Figure 7 (see Additional File 7 for an alternative representation). This model can be appreciated by considering the present-day human locus and then working backward in time using highly similar local sequences within the A3 locus, which provide 'footprints' for recent recombination events. First, full-length A3A and the Z1 domain of A3B are 97% identical, and they are flanked by nearly homologous ~5.5 kb regions (*i.e.*, direct repeats of 95% identity). These footprints strongly suggested that a recent duplication of two consecutive ancestral domains (Z1-Z2) gave rise to present-day A3B (event 7). Second, we inferred that this recent duplication resulted in a vestigial Z2 domain upstream of A3C, which was subsequently deleted prior to the divergence of human and chimpanzee lineages (event 8). Such a deletion event was supported by the fact that ~3 kb regions of 92% identical DNA reside upstream of the present-day A3B and A3C Z2

domains (these repeats lack similarity to other DNA within the locus). Third, a 92% similarity between two regions (~10 kb) encompassing the A3DE and A3F genes suggested they originated from a recent duplication. Moreover, a similar level of identity was found between two other regions (~10 kb) encompassing the Z2 domains of A3F and A3G. This strongly supported a common ancestral origin for the N-terminal domains of the A3DE, A3F and A3G genes (events 5 and 6). The likelihood of these four relatively recent events suggested that the ancestral locus configuration prior to event 5 [Z1-(Z2)₃-Z1-Z3] was a key intermediate in the evolution of the primate A3 locus (event 4 product in Figure 7).

Unequal crossing-over events prior to the ancestral intermediate were harder to infer because the footprints have been erased by sequence divergence. We therefore developed an algorithm to compute the minimal series of

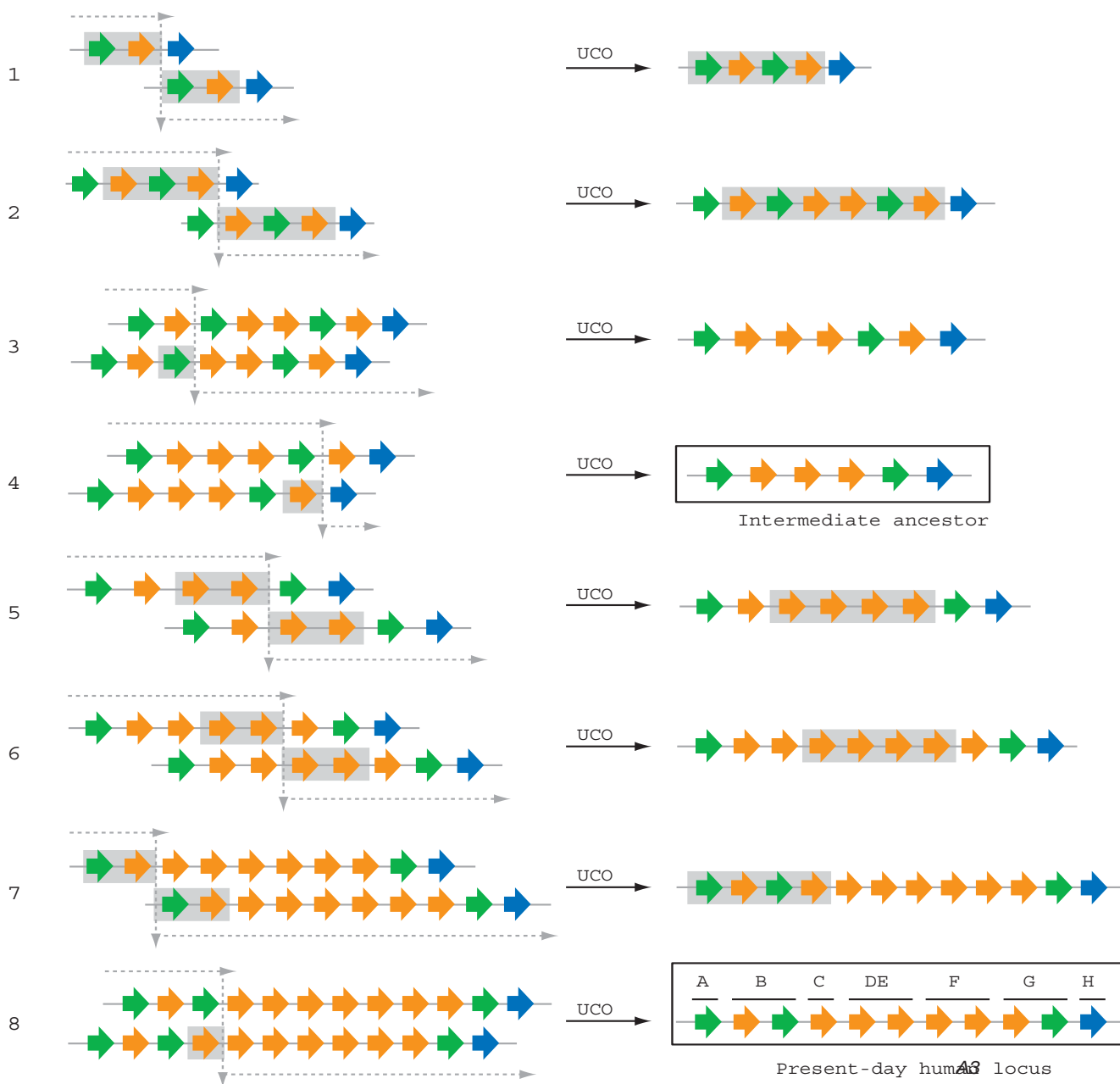


Figure 7
An 8-event model for the duplication and deletion history of the human A3 repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. Five duplication and three deletion events were predicted to transform the ancestral locus into the present-day human A3 repertoire. The first event was predicted to occur between two copies of the ancestral Z1-Z2-Z3 locus. The Z domain(s) affected by each unequal crossing-over (UCO) event is shaded gray. The crossing-over points are indicated by a dashed line arrows, and the resulting Z domain configurations are shown (we assumed that new configurations achieved homozygosity prior to being involved in a subsequent UCO). Although deletion events 3 and 4 are illustrated as interchromosomal UCOs, they could have also been caused by intrachromosomal events. Event 4 is depicted before an inferred 'intermediate ancestor' common to nearly all of our models and therefore considered parsimonious, but this event could have occurred any time after event 2. The underlying phylogeny for this model is identical to that shown in Figure 1, except the N-terminal domain of human A3B diverged prior to the point at which the N-terminal domains of human A3F/A3DE and A3G split. An alternative depiction of this model is shown in Additional File 7 and details can be found in the main text and Methods.

duplication and deletion events that could have generated this intermediate locus from the Z1-Z2-Z3 ancestor. Three minimal scenarios were found and each involved 4 events. However, when phylogenetic data were considered, only one scenario was plausible and it involved a 2-domain duplication, a 3-domain duplication and two single domain deletions (respectively, events 1 to 4 in Figure 7 & Additional File 7). Thus, together with the events detailed above, we inferred that the current human A3 repertoire is the product of 8 recombination events – 5 duplications and 3 deletions.

Theoretically, models with as few as 5 events are possible if the likely intermediate locus configuration is ignored. However, these models are also untenable as they clash with phylogenetic and local sequence alignment data. It should be noted that 8 events represent only a lower bound to explain the evolution of the A3 human locus. Scenarios involving more than 8 events could also lead to the same domain organization, and some events may have left no observable trace in the human lineage. Thus, this lower bound could increase when the complete A3 locus sequence of more mammals, and especially more primates, comes available. Finally, it is worth emphasizing that most (if not all) of the 8 recombination events modeled here happened in the 65 to 85 million year period between the points when the rodent and Old World monkey (*e.g.*, rhesus macaque) lineages split from the phylogenetic branch that led to humans (the time frame indicated by the asterisk in Figure 6).

Discussion

The present studies were initiated to gain a better understanding of the full A3 repertoire of three artiodactyl lineages – cattle, pigs and sheep – and to achieve insights into the mechanism and timing of the A3 gene expansion in mammals. We demonstrated that sheep and cattle have three A3 genes, A3Z1 A3Z2 and A3Z3. However, the latter two genes and their counterpart in pigs have the unique ability to produce a double-domain protein A3Z2-Z3, in addition to single-domain polypeptides. Thus, the A3 proteome of these species is more formidable than gene number alone would indicate. Our studies also help highlight the important point that, although A3 proteins consist of either one or two conserved Z domains, each of these domains can function and evolve independently.

Prior to the present studies, it was clear that most (if not all) placental mammals had Z2- and Z3-type A3 domains (*e.g.*, human, mouse, cat, pig, sheep and cow [35,36,42,54,72]). It was far less clear how broadly the Z1 domain distributed. Here, we presented two critical lines of evidence strongly indicating that the Z1 distribution is equally broad and, importantly, that the common ancestor of placental mammals had a Z1-Z2-Z3 A3 gene reper-

toire, similar to that of present-day sheep and cattle. First, the sheep and cattle A3 genomic sequences demonstrated the presence of a Z1-type A3 gene outside of the primate phylogenetic branches (Figure 6). Second, our pan-species Z1 PCR data, public EST data and draft genomic sequences from horses and dogs combined to show that a A3Z1 gene exists in other parts of the artiodactyl-containing phylogenetic branch set. These data supported a model in which the common ancestor of the primate- and the artiodactyl-containing mammalian super-orders, Euarchotheria and Laurasiatheria, respectively, had a A3Z1 gene and precisely one of each of the three conserved Z domain types (*i.e.*, a divergent model for A3 gene evolution, as opposed to one in which A3Z1 genes evolved independently in several limbs of the mammalian tree). We have therefore established a critical foundation for understanding the function(s) and evolutionary history of the A3 repertoire of any other placental mammal.

It is noteworthy that our pan-species Z1 PCR analyses failed to generate product from opossum genomic DNA and that the recently released opossum and platypus genomic sequences lack A3 genes (Figure 2C; [73,74]). This is unlikely to be a gap in the DNA sequence assemblies because, like non-mammalian vertebrates, DNA and protein searches clearly revealed the A3-flanking genes CBX6 and CBX7 in both animals (LaRue & Harris, unpublished data). Thus, unfortunately, these two interesting non-placental mammals are unlikely to provide significant insights into the earliest stages of A3 gene evolution (*i.e.*, pre-dating the Z1-Z2-Z3 ancestor described here). Perhaps data from the other two placental mammal super-orders, Afrotheria and Xenarthra (*e.g.*, represented by animals such as aardvarks and anteaters, respectively), will help shed light on earlier stages of A3 gene evolution, when presumably an AID-like gene transposed between CBX6 and CBX7 and duplicated to give rise to the ancestral Z1-Z2-Z3 locus. Nevertheless, because all current data indicate that the A3 genes are specific to placental mammals, we hypothesize that a unique role of these genes may relate to the placenta itself, where the A3 proteins may function to help protect the developing fetus from potentially harmful retrotransposition events and/or retroviral infections.

A growing body of evidence indicates that the sole function of the A3 genes of mammals is to provide an innate immune defense to retrovirus and retrotransposon mobilization. This is supported by the fact that the single A3 gene of mice is dispensable and that many of the mammalian A3 genes show evidence for a strong diversifying selection ([10,41,65,66] and this study, Table 1). Although the reason(s) are presently unknown, a large A3 repertoire is clearly more important for some mammals than it is for others. Humans, chimpanzees and rhesus

macaques have 11 Z domains, approximately 3- to 4-fold more than any other known non-primate mammal (Figure 6). Indeed, our studies indicated that the ancestors of humans and chimpanzees experienced at least eight Z domain recombination events, which is more than the total combined number of events for other known mammals. Therefore, despite the fact that the artiodactyl A3 genes show evidence for positive selection, their relative stability in copy number suggests that a considerable disadvantage – such as the potential to mutate genomic DNA – may outweigh the innate benefit of having numerous A3s to combat potentially invasive retroelements. This possibility may very well relate directly to an emerging trend in mammals, which is the frequent loss of a *A3Z1* gene which encodes a protein that can penetrate the nuclear compartment (*e.g.*, Figure 5). An *A3Z1* deletion was shown here for pigs, inferred here for cats and mice/rats, and demonstrated recently for some human populations (Figure 6 and [75]).

Finally, a major question is what selective pressure(s) drove the A3 expansion from an ancestral Z1-Z2-Z3 repertoire to the present day human Z1-Z2-Z1-(Z2)₆-Z1-Z3 repertoire? We propose that large-scale events such as gene expansions were selected by extremely pathogenic or lethal retroviral epidemics, because rare expansions would have been easily lost amongst a population of non-expanded alleles. A powerful selective pressure such as a lethal epidemic has the potential to produce a population bottleneck such that mostly (or only) pathogen-resistant individuals would survive (*i.e.*, those with the appropriate disease-resistant A3 repertoire). Such powerful selective pressures would have the potential to promote and perhaps even cause speciation events. We further predict that such events may be marked by changes in A3 Z domain copy number. It is therefore quite plausible that at least some of the eight recombination events required to transform the ancestral Z1-Z2-Z3 repertoire into the present day human Z1-Z2-Z1-(Z2)₆-Z1-Z3 repertoire may have protected our human ancestors from ancient retroviral infections and thereby facilitated the evolution of primates (a process that we have termed primatification).

Conclusion

The A3 locus of sheep and cattle consists of three genes, *A3Z1*, *A3Z2* and *A3Z3*, and the potential to encode four functional proteins, three directly and one (*A3Z2*-Z3) by read-through transcription and alternative splicing. The A3 locus of pigs experienced a deletion and therefore lacks *A3Z1*. The artiodactyl A3 repertoire demonstrates a unique modularity centered upon the conserved zinc-coordinating motifs. DNA deaminase activity data and subcellular localization studies suggest that this modularity may also correspond to a broader functionality. All of

the data combined to indicate that the common ancestor of artiodactyls and humans possessed a sheep/cattle-like A3 gene set, with the organization and capacity to evolve into the present day repertoires. The remarkable A3 gene expansion in the primate lineage – from the three ancestral genes (*A3Z1*-*A3Z2*-*A3Z3*) to the present-day eleven Z-domain human repertoire – was predicted to require a minimum of eight recombination events, most of which may have been required to thwart an ancient retroviral infections.

Methods

Genomic DNA sequences

A combination of array hybridization, A3-, *CBX6*- and *CBX7*-specific PCR was used to identify one A3-positive BACs for sheep (CHORI-243 clone 268D23; a kind gift from P. de Jong, BACPAC Resources Center, <http://bacpac.chori.org/library.php?id=162>) and two for pigs (RPCI-44 clones 344O17 and 408D3; [76]). *E. coli* were transformed with these BACs, grown to saturation in 50 ml cultures and used for DNA preparations as recommended (Marligen Biosciences). Purified BAC DNA was sheared to an average of approximately 3000 bp (Hydroshear method, Genomic Solutions). Fragment ends were blunted with T4 and Klenow DNA polymerases (NEB) and ligated into pBluescriptSK- (Stratagene) or pSMART-HC (Lucigen). Individual subclones were picked randomly and sequenced (ABI3730; Applied Biosystems). Phrap (P. Green, 1996, <http://www.phrap.org/phred/phrap/phrap.html>) and Sequencher 4.8 (Gene Codes Corp.) were used to assemble DNA sequences and they were groomed manually. Sequence coverage for the sheep A3 locus averaged 4.5 sequences and the pig 27 sequences. The genomic sequences were compared using Jdotter software (<http://www.jxxi.com/webstart/app/jdotter-a-java-dot-plot-viewer.jsp>; [77]). Repetitive sequences were identified using RepeatMasker <http://www.repeatmasker.org>.

A3 exons were identified by directly comparing the genomic DNA sequences with cDNA, EST and RACE sequences (below, Additional Files 3 & 8 and [42]). Predicted ISREs were identified and compared using the TransFac and Biobase databases through the softberry NSITE portal <http://www.softberry.com>. The sheep *CBX6* exons were identified with the help of GenBank EST sequences [EE808826.1](#), [DY519385.1](#) and [EE822736.1](#). The pig *CBX6* exons were also identified in this manner using [BP158234.1](#), [BP997823.1](#) & [BP153834.1](#). The sheep and pig *CBX7* exons were identified by homology to the cow gene (below). Other *CBX6* and *CBX7*, sequences, respectively, were [NM_014292.3](#) and [NM_175709.2](#) (human), [NM_001103094](#) and [XM_604126](#) (cow), [NM_028763.3](#) and [NM_144811](#) (mouse) and [NM_001016617.2](#) & [NM_001005071](#) (frog).

A3Z1 gene degenerate PCR analyses

Genomic DNA was isolated from the following tissues or cell lines: opossum kidney tissue, mouse NIH-3T3 cells, pig PK-15 cells, peccary brain tissue, cow MDBK cells, sheep FLK cells, horse blood cells (PBMC), African green monkey COS7 cells and human 293T cells (DNeasy, Qiagen). 10ng genomic DNA was used as template for PCR using primers designed to anneal to all known A3Z1 genes: 5'-GCC ATG CRG AGC TSY RCT TCY TGG and 5'-GTC ATD ATK GWR AYT YKG GCC CCA GC-3'. Two PCR rounds were used to achieve the final number of cycles (30 plus 18, 21 or 24 cycles). Amplicons were analyzed by agarose gel electrophoresis, TOPO-cloned (Invitrogen) and subjected to DNA sequencing. In all instances, the expected A3Z1 fragments were recovered (*e.g.*, Z1 of human A3A, A3B and A3G could all be detected in a single reaction). 30 PCR cycles using identical conditions and degenerate primers for the ALDOA gene were used as a positive control (5'-CGC TGT GCC CAG TAY AAG AAG GAY GG-3' and 5'-CTG CTG GCA RAT RCT GGC YTA).

Identifying expressed mRNAs by RACE

RNA was extracted from fresh pig (*Sus scrofa* Landrace/Yorkshire cross), sheep (*Ovis aries* Hampshire) and cattle (*Bos taurus* Hereford) PBMCs using the QIAamp RNA Blood mini kit (Qiagen). 5' and 3' RACE was performed using reagents from the FirstChoice RLM-RACE kit (Ambion). The protocol was modified slightly by using SAP (Roche) instead of CIP to remove 5'-phosphates. A3 cDNA 5' and 3' ends were amplified using Phusion high-fidelity polymerase (NEB), purified and TOPO-cloned (Invitrogen). All A3-specific primers used in conjunction with the 5' and 3' RACE primers are listed in Additional File 8.

A3 expression plasmids

The pTrc99A-based *E. coli* expression plasmids for sheep, cattle and pig A3Z2-Z3 and for human A3C and A3H were reported previously [42,56]. Other pTrc99A-based constructs were made by ligating KpnI- and SalI-digested PCR fragments into a similarly cut vector. Cow A3Z2 and A3Z3 were amplified from PBMC cDNA (above) using primers 5'-NNN NGA GCT CAG GTA CCA CCA TGC AAC CAG CCT ACC GAG GC & 5'-NNN NGT CGA CTC ACC CGA GAA TGT CCT C and 5'-NNN NGA GCT CAG GTA CCA CCA TGA CCG AGG GCT GGG C & 5'-NNN NGT CGA CCT AAA TTG GGG CCG TTA GGA T, respectively. Pig A3Z2 was amplified from the USMARC1 cDNA library [78] using primers 5'-NNN NGA GCT CAG GTA CCA CCA TGG ATC CTC AGC GCC TGA GAC and 5'-NNN NGT CGA CTC AGC GGT AAC AAA TCC.

Cow A3Z1 was a special case (see main text). It was amplified from PBMC cDNA (above) using primers 5'-NNN NGA GCT CAG GTA CCA C CA TGG ACG AAT ATA CCT

TCA CT and 5'-NNN NGT CGA CGT TTT GCT GAG TCT TGA G and TOPO-cloned into pCR-BLUNT-II-TOPO (Invitrogen). As a control, human A3A was amplified using 5'-NNN NGA GCT CGG TAC CAC CAT GGA AGC CAG CCC AGC and 5'-NNN NGT CGA CCC CAT CCT TCA GTT TCC CTG ATT CTG GAG and TOPO-cloned. Catalytic mutant derivatives of the cow A3Z1 and human A3A plasmids were constructed by site-directed mutagenesis (Stratagene) using oligonucleotides 5'-CCT GCC ATG CAG CGC TCT ACT TCC TG & 5'-CAG GAA GTA GAG CGC TGC ATG GCA GG and 5'-GGC CGC CAT GCG GCG CTG CGT TCT TG & 5'-CAA GAA GCG CAG CGC CGC ATG GCG GCC, respectively.

The artiodactyl A3 proteins were expressed in HeLa cells as N-terminal fusions to eGFP (pEGFP-N3; Clontech). Cow and pig A3Z2-Z3-eGFP and the human A3A-, A3C-, A3F- and A3H-eGFP constructs were reported previously [42,79]. Cow and pig A3Z2-eGFP plasmids were made by cloning SacI/SalI-digested PCR products generated using primers 5'-NNN NGA GCT CAG GTA CCA CCA TGC AAC CAG CCT ACC GAG GC & 5'-NNN NGT CGA CCC CGA GAA TGT CCT CAA G and 5'-NNN NGA GCT CAG GTA CCA CCA TGG ATC CTC AGC GCC TGA GAC & 5'-NNN NGT CGA CCC ACC TGG CGT GAG CAC C, respectively. Cow and pig A3Z3-eGFP plasmids were made similarly using primers 5'-NNN NGA GCT CAG GTA CCA CCA TGA CCG AGG GCT GGG C & 5'-NNN GTC GAC TCC AAT TGG GGC CGT TAG GAT and 5'-NNN NGA GCT CAG GTA CCA CCA TGA CCG AGG GCT GGG CT & 5'-NNN GTC GAC TCC TCT CGA GTC ACT TCT TGA, respectively.

Due to the toxicity of cow A3Z1 in *E. coli*, an A3Z1::intron-eGFP plasmid was made by overlapping PCR to join 3 separate fragments: A3Z1 exons 1 and 2 (primers 5'-NNN NGA GCT CAG GTA CCA C CA TGG ACG AAT ATA CCT TCA CT and 5'-CCT GGA CTC ACC TTG TTG CGC), an L1-derived intron ([80]; primers 5'-GTG AGT CCA GGA GAT GTT TCA and 5'-CTG TTG AGA TGA AAG GAG ACA) and A3Z1 exons 3-5 (primers 5'-CAT CTC AAC AGG GTT TGG ATC A and 5'-NNN NGT CGA CGT TTT GCT GAG TCT TGA G). The resulting PCR amplicon was digested with EcoRI and SalI and then ligated into a similarly cut pEGFP-N3 (Clontech).

Rif^R DNA deamination assays

Cytosine deaminase activity of the artiodactyl A3 protein variants was measured by quantifying the accumulation of Rif^R mutants in *ung*-deficient *E. coli* (*e.g.*, [42,56]). All A3 proteins were expressed from pTrc99A (AP Biotech), with the exception of cow A3Z1 and human A3A, which were expressed using pCR-BLUNT-II-TOPO (Invitrogen). Experiments were done a minimum of three times, in the presence or absence of IPTG as indicated.

Fluorescence microscopy

To observe subcellular localization of A3 proteins, 5000 HeLa cells were incubated for 24 h in Labtek chambered coverglasses (Nunc), transfected with 200 ng of the pEGFP-N3 based constructs and, after an additional 24 h visualized on a Zeiss Axiovert 200 microscope at 400× magnification. HsA3F, HsA3C, HsA3H, HsA3A, BtA3Z2-Z3 and SsA3Z2-Z3 fusion constructs were previously reported [19,42,79].

Phylogenies and positive selection calculations

Z domain exons were used for all phylogenetic, positive selection and modelling studies. GARD showed no evidence for recombination breakpoints within the Z domain exons [81]. T_coffee version 5.31 was used for multiple sequence alignments [82]. PAL2NAL software was used to convert amino acid sequences to nucleotides [83]. JalView was used to remove insertions/deletions [84]. The dnaml program within the Phylip software package was used to generate a phylogenetic tree ([85]; an identical tree was obtained with MrBayes version 3.1 [86], except branch lengths differed slightly). Clustal W version 1.83.1 was also used for some individual domain comparisons [87].

Free ratio model positive selection studies were based on a phylogenetic tree generated through Bayesian inference using MrBayes version 3.1 [86]. Each tree was run for 250,000 generations with a burnin of 62,500 and standard default parameters. The PAML codeml program [88] was used to generate dN/dS ratios (ω values) for phylogenetic tree branches. ω values from the free ratio model using the F3 × 4 algorithm are shown in Additional File 6 (values from the F1 × 61 algorithm were similar and therefore not shown).

Positive selection was also evaluated in specific phylogenetic lineages using the NsSites model in the PAML codeml program (Table 1). Individual Z domain phylogenetic trees were generated as described above and used in these analyses. Z2 and Z3 comparisons were done for sheep, cow, pig, peccary and horse sequences, and Z1 comparisons for sheep, cow, horse and dog sequences (non-artiodactyl sequences were added for statistical significance; GenBank accession numbers are in Additional File 3). Models for neutral selection (M1 and M7) were compared to those for positive selection (M2 and M8). Likelihood ratio tests were performed to compare the null and positive selection scenarios.

A3 gene expansion modelling

The aim was to infer the most likely histories of duplications and deletions that gave rise to the human A3 locus. Instead of considering each gene as an individual element,

we subdivided it into its N-terminal and C-terminal Z-domains. Hence, the present-day human locus configuration was represented as follows: Z1-Z2-Z1-(Z2)₆-Z1-Z3. The considered duplications are 'multiple tandem duplications' resulting from unequal crossing over [70]. In other words, a single duplication event can copy an arbitrary number of consecutive Z-domains, and place them in the same order next to the original ones. Similarly, an unequal crossing-over can remove an arbitrary number of adjacent domains and cause deletions.

Various algorithms have been proposed to infer evolutionary histories of tandemly arrayed gene families [71,89-91], but none of them involve both multiple tandem duplications and deletions. Consequently, we developed a brute force algorithm to enumerate all possible evolutionary scenarios involving a minimum number of duplications and deletions that can transform a particular locus configuration into another. Such an exhaustive algorithm has an exponential time complexity and it is impractical for analyzing large gene families. However, the limited size of the A3 locus and the classification of the Z domains into three distinct categories made it useful here (e.g., events 1 to 4 in Figure 7 & Additional File 7).

To infer the most recent evolutionary events (events 5 to 8 in Figure 7 & Additional File 7), we performed an analysis of the self-similarities within the human A3 locus. The DNA sequence (hg18, chr22:37682569-37830946) with identified interspersed repeats was downloaded from the RepeatMasker web site <http://www.repeatmasker.org>. A dot plot of this sequence with itself was obtained using Gepard [92] to identify pairs of regions with very high similarities. The three most significant were extracted and further aligned using Blastz [93] with default parameter to obtain the percentage of identity. These regions were used to infer and model the most recent evolutionary events, as described in the main text.

Data deposition

The GenBank accession number for the sheep A3 genomic sequence is [FI042940](http://www.ncbi.nlm.nih.gov/GenBank/FASTA/FI042940). The GenBank accession numbers for the two pig A3 genomic sequences are [FI042938](http://www.ncbi.nlm.nih.gov/GenBank/FASTA/FI042938) and [FI042939](http://www.ncbi.nlm.nih.gov/GenBank/FASTA/FI042939). All A3 cDNA and EST sequences have also been deposited (see Additional File 3 for a full list of GenBank accession numbers).

Abbreviations

A.A.: amino acid; A3: APOBEC3; A3A: APOBEC3B; A3B: APOBEC3C; A3C: APOBEC3C; A3DE: APOBEC3DE; A3F: APOBEC3F; A3G: APOBEC3G; A3H: APOBEC3H; GFP: Green Fluorescent Protein; PAML: Phylogenetic Analysis by Maximum Likelihood; MYA: Millions of Years Ago; Z: Zinc-coordinating motif

Authors' contributions

RSL and RSH designed the studies, performed experiments, analyzed data and wrote the manuscript. SRJ and VA helped analyze the artiodactyl A3 genes and proteins, TPLS provided library samples and generated genomic DNA sequences, IH contributed cattle A3 gene sequences and functional data, KATS assisted with phylogenetic and computational studies, and ML, DB and NE generated the model for A3 evolution. All authors contributed to editing the manuscript.

Additional material

Additional file 1

APOBEC3 Z domain conservation. Web LOGO profiles depicting amino acid conservation within each mammalian Z domain. The multiple sequence alignments used to generate the phylogenetic tree in Figure 1 were used to create consensus profiles for each of the indicated Z domains using Web LOGO [94]. Arrowheads below the amino acid profiles indicate residues that define each Z type (see the main text for additional details).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S1.ps>]

Additional file 2

APOBEC3 genomic locus comparisons. Dotplot alignments of (A) the sheep and pig and (B) the sheep and cow A3 genomic sequences.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S2.eps>]

Additional file 3

Mammalian A3 and AID sequences. A table summarizing all DNA sequences used in this study, including GenBank accession numbers.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S3.eps>]

Additional file 4

APOBEC3 promoter element conservation. Predicted interferon-stimulating response elements (ISRE) in the promoter regions of the indicated A3 genes and known interferon-inducing genes ISG54 and ISG15. The ISRE sequences are shown relative to the translation initiation codon ATG. Identities to human sequences are shaded gray.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S4.eps>]

Additional file 5

E. coli-based DNA cytosine deaminase activity data. DNA cytosine deaminase activity of the pig A3Z2-Z3 and A3Z2 proteins in E. coli. Conditions and labels are identical to those used in Figure 4, except 10 independent cultures were grown under IPTG-induced conditions and analyzed.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S5.eps>]

Additional file 6

Evidence for positive selection in APOBEC3 gene evolution. Phylogenetic trees showing relative relationships and ω values for the indicated (A) Z1 domains, (B) Z2 domains, (C) Z3 domains and (D) the Z domain of AID. The phylogenetic trees were determined using MrBayes, and the ω values were calculated using the PAML free ratio model. ω values are shown in red adjacent to (or where space is non-permitting, to right of) each phylogenetic branch. Asterisks denote branches where the ω value was infinity (i.e., dS was zero). The units for the scale bars are nucleotide changes per codon. The dotted line in panel (C) was used to provide more space to depict the human and non-human primate Z3 tree branches. See the main text and Methods for additional details.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S6.eps>]

Additional file 7

Proposed APOBEC3 gene diversification events during primatification. An alternative representation of the 8-event model for the duplication and deletion history of the human A3 repertoire. Z1, Z2 and Z3 domains are colored green, orange and blue, respectively. The Z domain(s) involved in each event are shaded gray. Dark black and red lines mark duplications (one color for the original segment and one color for the duplicated segment), crosses designate deletions and light gray lines indicate no change. See the main text, Figure 7 and Methods for details.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S7.doc>]

Additional file 8

Primers used to identify expressed APOBEC3 transcripts from cow, sheep and pig PBMCs. A table summarizing the oligonucleotide primers used in this study.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2199-9-104-S8.doc>]

Acknowledgements

We thank L. Beach, H. Malik, M. Murtaugh and M. Stenglein for valuable feedback. We thank K. Tennill and R. Godtel for assistance with BAC DNA sequencing, D. Shiroma and S. Fahrenkrug for help identifying pig BAC clones, M. Stenglein for several expression plasmids, M. Titus for use of her microscope, L. Hartman for pig and sheep samples, C. Knutson for cow samples, J. Zimmerman and R. Molina for peccary blood, P. Krauseman for peccary brain tissue and M. Ruen and O. Holland for opossum samples. R. LaRue is a member of the University of Minnesota CMB Graduate Program. S. Jónsson was the 2004–2005 Val Bjornson Icelandic Exchange Scholarship recipient. M. Lajoie was supported by a Canadian Institutes of Health Research studentship. D. Bertrand and N. El-Mabrouk were supported by grants from the Fonds Québécois de la Recherche sur la Nature et les Technologies and the Natural Sciences and Engineering Research Council of Canada. R. Harris was supported in part by a Searle Scholarship and a University of Minnesota McKnight Land Grant Assistant Professorship. This work was also supported by NIH grant AI064046. The University of Minnesota Advanced Genetic Analysis Facility assisted with DNA sequencing.

References

1. Bieniasz PD: **Intrinsic immunity: a front-line defense against viral attack.** *Nat Immunol* 2004, **5**(11):1109-1115.

2. Chiu YL, Greene WC: **The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements.** *Annu Rev Immunol* 2008, **26**:317-353.
3. Cullen BR: **Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors.** *J Virol* 2006, **80**(3):1067-1076.
4. Goff SP: **Retrovirus restriction factors.** *Mol Cell* 2004, **16**(6):849-859.
5. Malim MH, Emerman M: **HIV-1 accessory proteins – ensuring viral survival in a hostile environment.** *Cell Host Microbe* 2008, **3**(6):388-398.
6. Doehle BP, Bogerd HP, Wiegand HL, Jouvenet N, Bieniasz PD, Hunter E, Cullen BR: **The betaretrovirus Mason-Pfizer monkey virus selectively excludes simian APOBEC3G from virion particles.** *J Virol* 2006, **80**(24):12102-12108.
7. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, Neuberger MS, Malim MH: **DNA deamination mediates innate immunity to retroviral infection.** *Cell* 2003, **113**(6):803-809.
8. Jost S, Turelli P, Mangeat B, Protzer U, Trono D: **Induction of antiviral cytidine deaminases does not explain the inhibition of hepatitis B virus replication by interferons.** *J Virol* 2007, **81**(19):10588-10596.
9. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D: **Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts.** *Nature* 2003, **424**(6944):99-103.
10. Okeoma CM, Lovsin N, Peterlin BM, Ross SR: **APOBEC3 inhibits mouse mammary tumour virus replication in vivo.** *Nature* 2007, **445**(7130):927-930.
11. Suspene R, Guetard D, Henry M, Sommer P, Wain-Hobson S, Vartanian JP: **Extensive editing of both hepatitis B virus DNA strands by APOBEC3 cytidine deaminases in vitro and in vivo.** *Proc Natl Acad Sci USA* 2005, **102**(23):8321-8326.
12. Turelli P, Mangeat B, Jost S, Vianin S, Trono D: **Inhibition of hepatitis B virus replication by APOBEC3G.** *Science* 2004, **303**(5665):1829.
13. Wiegand HL, Cullen BR: **Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins.** *J Virol* 2007, **81**(24):13694-13699.
14. Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L: **The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA.** *Nature* 2003, **424**(6944):94-98.
15. Armitage AE, Katzourakis A, de Oliveira T, Welch JJ, Belshaw R, Bishop KN, Kramer B, McMichael AJ, Rambaut A, Iversen AK: **Conserved footprints of APOBEC3G on hypermutated HIV-1 and HERV-K(HML2) sequences.** *J Virol* 2008, **82**(17):8743-8761.
16. Dutko JA, Schafer A, Kenny AE, Cullen BR, Curcio MJ: **Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases.** *Curr Biol* 2005, **15**(7):661-666.
17. Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, Hance AJ, Heidmann T, Schwartz O: **APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses.** *Nature* 2005, **433**(7024):430-433.
18. Jern P, Stoye JP, Coffin JM: **Role of APOBEC3 in genetic diversity among endogenous murine leukemia viruses.** *PLoS Genet* 2007, **3**(10):2014-2022.
19. Jónsson SR, LaRue RS, Stenglein MD, Fahrenkrug SC, Andrésdóttir V, Harris RS: **The restriction of zoonotic PERV transmission by human APOBEC3G.** *PLoS ONE* 2007, **2**(9):e893.
20. Kaiser SM, Malik HS, Emerman M: **Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein.** *Science* 2007, **316**(5832):1756-1758.
21. Lee YN, Bieniasz PD: **Reconstitution of an infectious human endogenous retrovirus.** *PLoS Pathog* 2007, **3**(1):e10.
22. Lee YN, Malim MH, Bieniasz PD: **Hypermutation of an ancient human retrovirus by APOBEC3G.** *J Virol* 2008, **17**:8762-8770.
23. Schumacher AJ, Nissley DJ, Harris RS: **APOBEC3G hypermutates genomic DNA and inhibits Ty1 retrotransposition in yeast.** *Proc Natl Acad Sci USA* 2005, **102**(28):9854-9859.
24. Chen H, Lilley CE, Yu Q, Lee DV, Chou J, Narvaiza I, Landau NR, Weitzman MD: **APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons.** *Curr Biol* 2006, **16**(5):480-485.
25. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T, Sata T, Tokunaga K: **All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition.** *Nucleic Acids Res* 2007, **35**(9):2955-2964.
26. Turelli P, Vianin S, Trono D: **The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay.** *J Biol Chem* 2004, **279**(42):43371-43373.
27. Muckenfuss H, Hamdorf M, Held U, Perkovic M, Lower J, Cichutek K, Flory E, Schumann GG, Munk C: **APOBEC3 proteins inhibit human LINE-1 retrotransposition.** *J Biol Chem* 2006, **281**(31):22161-22172.
28. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR: **Cellular inhibitors of long interspersed element 1 and Alu retrotransposition.** *Proc Natl Acad Sci USA* 2006, **103**(23):8780-8785.
29. Niewiadomska AM, Tian C, Tan L, Wang T, Thi Nguyen Sarkis P, Yu XF: **Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with HMM complex formation or P-body association.** *J Virol* 2007, **17**:9577-9583.
30. Stenglein MD, Harris RS: **APOBEC3B and APOBEC3F inhibit LI retrotransposition by a DNA deamination-independent mechanism.** *J Biol Chem* 2006, **281**(25):16837-16841.
31. Abudu A, Takaori-Kondo A, Izumi T, Shirakawa K, Kobayashi M, Sasada A, Fukunaga K, Uchiyama T: **Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms.** *Curr Biol* 2006, **16**(15):1565-1570.
32. Derse D, Hill SA, Princler G, Lloyd P, Heidecker G: **Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid.** *Proc Natl Acad Sci USA* 2007, **104**(8):2915-2920.
33. Russell RA, Wiegand HL, Moore MD, Schafer A, McClure MO, Cullen BR: **Foamy virus Bet proteins function as novel inhibitors of the APOBEC3 family of innate antiretroviral defense factors.** *J Virol* 2005, **79**(14):8724-8731.
34. Yu X, Yu Y, Liu B, Luo K, Kong W, Mao P, Yu XF: **Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex.** *Science* 2003, **302**(5647):1056-1060.
35. Conticello SG, Thomas CJ, Petersen-Mahrt S, Neuberger MS: **Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases.** *Mol Biol Evol* 2005, **22**(2):367-377.
36. Harris RS, Liddament MT: **Retroviral restriction by APOBEC proteins.** *Nat Rev Immunol* 2004, **4**(11):868-877.
37. Larue RS, Andrésdóttir V, Blanchard Y, Conticello SG, Derse D, Emerman M, Greene WC, Jónsson SR, Landau NR, Löchelt M, et al.: **Guidelines for naming non-primate APOBEC3 genes and proteins.** *J Virol* 2008 in press.
38. Haché G, Liddament MT, Harris RS: **The retroviral hypermutation specificity of APOBEC3F and APOBEC3G is governed by the C-terminal DNA cytosine deaminase domain.** *J Biol Chem* 2005, **280**(12):10920-10924.
39. Navarro F, Bollman B, Chen H, König R, Yu Q, Chiles K, Landau NR: **Complementary function of the two catalytic domains of APOBEC3G.** *Virology* 2005, **333**(2):374-386.
40. Newman EN, Holmes RK, Craig HM, Klein KC, Lingappa JR, Malim MH, Sheehy AM: **Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity.** *Curr Biol* 2005, **15**(2):166-170.
41. OhAinle M, Kerns JA, Malik HS, Emerman M: **Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H.** *J Virol* 2006, **80**(8):3853-3862.
42. Jónsson SR, Haché G, Stenglein MD, Fahrenkrug SC, Andrésdóttir V, Harris RS: **Evolutionarily conserved and non-conserved retrovirus restriction activities of artiodactyl APOBEC3F proteins.** *Nucleic Acids Res* 2006, **34**(19):5683-5694.
43. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: **The delayed rise of present-day mammals.** *Nature* 2007, **446**(7135):507-512.
44. Kumar S, Hedges SB: **A molecular timescale for vertebrate evolution.** *Nature* 1998, **392**(6679):917-920.
45. Reza Shariflou M, Moran C: **Conservation within artiodactyls of an AATA interrupt in the IGF-1 microsatellite for 19–35 million years.** *Mol Biol Evol* 2000, **17**(4):665-669.
46. Peng G, Lei KJ, Jin W, Greenwell-Wild T, Wahl SM: **Induction of APOBEC3 family proteins, a defensive maneuver underlying**

- interferon-induced anti-HIV-1 activity. *J Exp Med* 2006, **203(1)**:41-46.
47. Tanaka Y, Marusawa H, Seno H, Matsumoto Y, Ueda Y, Kodama Y, Endo Y, Yamauchi J, Matsumoto T, Takaori-Kondo A, et al.: **Antiviral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes.** *Biochem Biophys Res Commun* 2006, **341(2)**:314-319.
 48. Abrahams VM, Schaefer TM, Fahey JV, Visintin I, Wright JA, Aldo PB, Romero R, Wira CR, Mor G: **Expression and secretion of antiviral factors by trophoblast cells following stimulation by the TLR-3 agonist, Poly(I : C).** *Hum Reprod* 2006, **21(9)**:2432-2439.
 49. Bonvin M, Achermann F, Greeve I, Stroka D, Keogh A, Inderbitzin D, Candinas D, Sommer P, Wain-Hobson S, Vartanian JP, et al.: **Interferon-inducible expression of APOBEC3 editing enzymes in human hepatocytes and inhibition of hepatitis B virus replication.** *Hepatology* 2006, **43(6)**:1364-1374.
 50. Kessler DS, Levy DE, Darnell JE Jr: **Two interferon-induced nuclear factors bind a single promoter element in interferon-stimulated genes.** *Proc Natl Acad Sci USA* 1988, **85(22)**:8521-8525.
 51. Levy DE, Kessler DS, Pine R, Reich N, Darnell JE Jr: **Interferon-induced nuclear factors that bind a shared promoter element correlate with positive and negative transcriptional control.** *Genes Dev* 1988, **2(4)**:383-393.
 52. Reich N, Evans B, Levy D, Fahey D, Knight E Jr, Darnell JE Jr: **Interferon-induced transcription of a gene encoding a 15-kDa protein depends on an upstream enhancer element.** *Proc Natl Acad Sci USA* 1987, **84(18)**:6394-6398.
 53. Muckenfuss H, Kaiser JK, Krebil E, Battenberg M, Schwer C, Cichutek K, Munk C, Flory E: **Sp1 and Sp3 regulate basal transcription of the human APOBEC3G gene.** *Nucleic Acids Res* 2007, **35(11)**:3784-3796.
 54. Münk C, Beck T, Zielonka J, Hotz-Wagenblatt A, Chareza S, Battenberg M, Thielebein J, Cichutek K, Bravo IG, O'Brien SJ, et al.: **Functions, structure, and read-through alternative splicing of feline APOBEC3 genes.** *Genome Biol* 2008, **9(3)**:R48.
 55. Dang Y, Wang X, Esselman WJ, Zheng YH: **Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family.** *J Virol* 2006, **80(21)**:10522-10533.
 56. Harris RS, Petersen-Mahrt SK, Neuberger MS: **RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators.** *Molecular Cell* 2002, **10(5)**:1247-1253.
 57. Liddament MT, Brown WL, Schumacher AJ, Harris RS: **APOBEC3F properties and hypermutation preferences indicate activity against HIV-1 in vivo.** *Curr Biol* 2004, **14(15)**:1385-1391.
 58. Mariani R, Chen D, Schröfelbauer B, Navarro F, König R, Bollman B, Münk C, Nymark-McMahon H, Landau NR: **Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif.** *Cell* 2003, **114(1)**:21-31.
 59. Yu Q, Chen D, König R, Mariani R, Unutmaz D, Landau NR: **APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication.** *J Biol Chem* 2004, **279(51)**:53379-53386.
 60. Iwatani Y, Takeuchi H, Strebler K, Levin JG: **Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect.** *J Virol* 2006, **80(12)**:5992-6002.
 61. Bogerd HP, Wiegand HL, Doehle BP, Cullen BR: **The intrinsic antiretroviral factor APOBEC3B contains two enzymatically active cytidine deaminase domains.** *Virology* 2007, **364(2)**:486-493.
 62. Dang Y, Siew LM, Wang X, Han Y, Lampen R, Zheng YH: **Human cytidine deaminase APOBEC3H restricts HIV-1 replication.** *J Biol Chem* 2008, **283(17)**:11606-11614.
 63. OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M: **Antiretroelement activity of APOBEC3H was lost twice in recent human evolution.** *Cell Host Microbe* 2008, **4(3)**:249-259.
 64. Chen KM, Martemyanova N, Lu Y, Shindo K, Matsuo H, Harris RS: **Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G.** *FEBS Lett* 2007, **581(24)**:4761-4766.
 65. Sawyer SL, Emerman M, Malik HS: **Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G.** *PLoS Biol* 2004, **2(9)**:E275.
 66. Zhang J, Webb DM: **Rapid evolution of primate antiviral enzyme APOBEC3G.** *Hum Mol Genet* 2004, **13(16)**:1785-1791.
 67. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, et al.: **Evolutionary and biomedical insights from the rhesus macaque genome.** *Science* 2007, **316(5822)**:222-234.
 68. Virgen CA, Hatzioannou T: **Antiretroviral activity and Vif sensitivity of rhesus macaque APOBEC3 proteins.** *J Virol* 2007, **81(24)**:13932-13937.
 69. Purvis A: **A composite estimate of primate phylogeny.** *Philos Trans R Soc Lond B Biol Sci* 1995, **348(1326)**:405-421.
 70. Fitch WM: **Phylogenies constrained by the crossover process as illustrated by human hemoglobins and a thirteen-cycle, eleven-amino-acid repeat in human apolipoprotein A-I.** *Genetics* 1977, **86(3)**:623-644.
 71. Gascuel O, Bertrand D, Elemento O: **Reconstructing the duplication history of tandemly repeated sequences.** *"Mathematics of Evolution and Phylogeny"* 2007:205-235.
 72. Jarmuz A, Chester A, Bayliss J, Gisbourne J, Dunham I, Scott J, Navaratnam N: **An anthropoid-specific locus of orphan C to U RNA-editing enzymes on chromosome 22.** *Genomics* 2002, **79(3)**:285-296.
 73. Mikkelsen TS, Wakefield MJ, Aken B, Amemiya CT, Chang JL, Duke S, Garber M, Gentles AJ, Goodstadt L, Heger A, et al.: **Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences.** *Nature* 2007, **447(7141)**:167-177.
 74. Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grutzner F, Belov K, Miller W, Clarke L, Chinwalla AT, et al.: **Genome analysis of the platypus reveals unique signatures of evolution.** *Nature* 2008, **453(7192)**:175-183.
 75. Kidd JM, Newman TL, Tuzun E, Kaul R, Eichler EE: **Population stratification of a common APOBEC gene deletion polymorphism.** *PLoS Genet* 2007, **3(4)**:e63.
 76. Fahrenkrug SC, Rohrer GA, Freking BA, Smith TP, Osoegawa K, Shu CL, Catanese JJ, de Jong PJ: **A porcine BAC library with tenfold genome coverage: a resource for physical and genetic map integration.** *Mamm Genome* 2001, **12(6)**:472-474.
 77. Brodie R, Smith AJ, Roper RL, Tcherepanov V, Upton C: **Base-By-Base: single nucleotide-level analysis of whole viral genome alignments.** *BMC Bioinformatics* 2004, **5**:96.
 78. Fahrenkrug SC, Smith TP, Freking BA, Cho J, White J, Vallet J, Wise T, Rohrer G, Perlea G, Sultana R, et al.: **Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly.** *Mamm Genome* 2002, **13(8)**:475-478.
 79. Stenglein MD, Matsuo H, Harris RS: **Two regions within the amino-terminal half of APOBEC3G cooperate to determine cytoplasmic localization.** *J Virol* 2008, **82(19)**:9591-9599.
 80. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr: **High frequency retrotransposition in cultured mammalian cells.** *Cell* 1996, **87(5)**:917-927.
 81. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelck CH, Frost SD: **GARD: a genetic algorithm for recombination detection.** *Bioinformatics* 2006, **22(24)**:3096-3098.
 82. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302(1)**:205-217.
 83. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.** *Nucleic Acids Res* 2006:V6609-612.
 84. Clamp M, Cuff J, Searle SM, Barton GJ: **The Jalview Java alignment editor.** *Bioinformatics* 2004, **20(3)**:426-427.
 85. Retief JD: **Phylogenetic analysis using PHYLIIP.** *Methods Mol Biol* 2000, **132**:243-258.
 86. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19(12)**:1572-1574.
 87. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al.: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23(21)**:2947-2948.
 88. Yang Z: **PAML: a program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13(5)**:555-556.
 89. Bertrand D, Lajoie M, El-Mabrouk N: **Inferring ancestral gene orders for a family of tandemly arrayed genes.** *J Comp Biol* 2008, **15(8)**:1063-1077.
 90. Lajoie M, Bertrand D, El-Mabrouk N, Gascuel O: **Duplication and inversion history of a tandemly repeated genes family.** *J Comput Biol* 2007, **14(4)**:462-478.

91. Zhang L, Ma B, Wang L, Xu Y: **Greedy method for inferring tandem duplication history.** *Bioinformatics* 2003, **19(12)**:1497-1504.
92. Krumsiek J, Arnold R, Rattei T: **Gepard: a rapid and sensitive tool for creating dotplots on genome scale.** *Bioinformatics* 2007, **23(8)**:1026-1028.
93. Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W: **Human-mouse alignments with BLASTZ.** *Genome Res* 2003, **13(1)**:103-107.
94. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14(6)**:1188-1190.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

