

1 An ancient endogenous DNA virus in the human genome

2

3 Jose Gabriel Nino Barreat,^a Aris Katzourakis,^{a#}

4

5 ^aDepartment of Biology, University of Oxford, Oxford, United Kingdom.

6

7 Running Head: Ancient endogenous DNA virus in the human genome

8

9 #Address correspondence to Aris Katzourakis, aris.katzourakis@zoo.ox.ac.uk

10

11 Abstract word count: 246/250

12 Importance word count: 146/150

13 Text word count: 2987 (excluding references, table footnotes and figure legends)

14

15

16

17

18

19

20

21

22

23

Abstract

The genomes of eukaryotes preserve a striking diversity of ancient viruses in the form of endogenous viral elements (EVEs). Study of this genomic fossil record provides insights into the diversity, origin and evolution of viruses across geological timescales. In particular, *Mavericks* have emerged as one of the oldest groups of viruses infecting vertebrates (≥ 419 My). They have been found in the genomes of fish, amphibians and non-avian reptiles but had been overlooked in mammals. Thus, their evolutionary history and the causes of their demise in mammals remain puzzling questions. Here, we conduct a detailed evolutionary study of two *Maverick*-like integrations found on human chromosomes 7 and 8. We performed a comparative analysis of the integrations and determined their orthology across placental mammals (Eutheria) via the syntenic arrangement of neighbouring genes. The integrations were absent at the orthologous sites in the genomes of marsupials and monotremes. These observations allowed us to reconstruct a time-calibrated phylogeny and infer the age of their most recent common ancestor at 268.61 (199.70–344.54) My. In addition, we estimate the age of the individual integrations at ~105 My which represent the oldest non-retroviral EVEs found in the human genome. Our findings suggest that active *Mavericks* existed in the ancestors of modern mammals ~172 My ago (Jurassic Period) and potentially to the end of the Early Cretaceous. We hypothesise *Mavericks* could have gone extinct in mammals from the evolution of an antiviral defence system or from reduced opportunities for transmission in terrestrial hosts.

Importance

The genomes of vertebrates preserve an enormous diversity of endogenous viral elements (remnants of ancient viruses that accumulate in host genomes over evolutionary time). Although retroviruses account for the vast majority of these elements, diverse DNA viruses have also been found and novel lineages are being described. Here we analyse two elements found in the human genome belonging to an ancient group of DNA viruses called *Mavericks*. We study their evolutionary history, finding that the elements are shared between humans and many different species of placental mammals. These observations suggest the elements inserted at least ~105 Mya in the most recent common ancestor of placentals. We further estimate the age of the viral ancestor around 268 My. Our results provide evidence for some of the oldest viral integrations in the human genome and insights into the ancient interactions of viruses with the ancestors of modern-day mammals.

Keywords

Comparative genomics, endogenous viral elements (EVEs), *Homo sapiens*, paleovirology.

Introduction

The genomes of vertebrates are teeming with viruses. Some of these elements are still active (they can replicate and produce viral progeny), while most are degraded genomic fossils. The most prominent are the retroviruses (RNA-retrotranscribing viruses) which can actively integrate and account for a significant proportion of host genomes (1). Remarkably, some animal DNA viruses also share this property of actively integrating into host genomes, including *Mavericks* (2, 3), *Teratorns* (4–6), herpesviruses (7–9) and adeno-associated virus (AAV) (10, 11).

Mavericks are a group of eukaryotic mobile genetic elements that encode a family B polymerase, a retroviral-like integrase and homologues of capsid morphogenesis genes found in dsDNA viruses (2, 3, 12, 13). Evolutionary analyses and comparisons of protein structure by homology modelling indicate that *Mavericks* are firmly nested within the viral kingdom *Bamfordvirae* (12–15). They are considered to be a lineage of endogenising DNA viruses that can undergo exogenous transmission (16, 17), although there is still no direct experimental evidence for their assembly into viral particles.

In vertebrates, *Mavericks* have been described in the genomes of teleost fish, coelacanths, amphibians, birds and non-avian reptiles. Most of the potentially active elements occur in teleost fish (97%) while only 4 intact copies (3%) have been found in tetrapod genomes (16). All the elements found in birds seem to be degraded and they had not been identified

in mammals (2, 3, 16). The reason for the deterioration and eventual loss of the elements from birds and mammals remains unclear.

Here we report the discovery of two novel placental-wide *Maverick* orthologues also found in humans which allow us to conduct a detailed evolutionary analysis of the elements across mammals (human sequences on chromosome 7 have been identified in previous work (17), but their orthology and evolution have not been investigated). By comparing these loci and placing them in a phylogenetic context, we gain new insights into the evolutionary history of *Mavericks* in mammals and the forces that have shaped their evolution. Furthermore, we discuss several hypotheses which could explain the demise of the elements in bird and mammalian genomes. The findings presented here illuminate our understanding about the evolutionary history of *Mavericks*, the endogenous viral elements in the human genome and the ancient interactions between viruses and our direct ancestors.

Methods

We used the protein-primed polymerase B (PolB) of the common box-turtle (*Terrapene carolina*) *Maverick* (16, 17) as a probe to screen the human genome with tBLASTn (18) (assembly GRCh38.p13, BLOSUM 45 substitution matrix, word size = 2). We obtained two hits with e-values < 1e-7: one on chromosome 7 (chr7:37601124–37601615, query cover = 15%, e-value = 8e-8, percent identity = 31.14%) and another on chromosome 8 (chr8:12858023–12858613, query cover = 25%, e-value = 5e-8, percent identity = 29.95%). In order to assess orthology of the integrations in the genomes of other mammals,

we identified the three most proximal protein-coding genes upstream and downstream of the integrations using the Ensembl genome browser (19).

The protein-coding genes used as genetic landmarks for the chromosome 7 integration were: *NME8*, *EPDR1*, *GPR141*, *ELMO1*, *AOAH* and *ANLN*, and for chromosome 8: *C8orf48*, *DLC1*, *TRMT9B*, *LONRF1*, *FAM86B2* and *DEFB130A* (NCBI protein database (20)). We used these sequences as queries in tBLASTn (set to the default) to screen the NCBI RefSeq genome database using the following taxon labels: ‘Monotremata (taxid: 9255)’, ‘Metatheria (taxid: 9263)’, ‘Afrotheria (taxid: 311790)’, ‘Xenarthra (taxid: 9348)’, ‘Laurasiatheria (taxid: 314145)’ and ‘Euarchontoglires (taxid: 314146)’. To verify the presence or absence of the integration, we screened the regions flanked by the genetic landmarks with the full set of proteins coded by the 4 tetrapod *Mavericks* reported previously (16) again with tBLASTn (e-value < 1e-5).

As an independent test for the presence of the integrations, we used the whole-genome alignment of 120 mammals published by Hecker and Hiller (21). We extracted the regions corresponding to the human *polB* integrations and extended them assuming the ancestral gene coded for 1053 amino acids as in the box-turtle *Maverick* (chr7:37600749–37603904, chr8:12856919–12860047). The regions were extracted from the chr7 and chr8 whole-genome alignments in MAF format (85.16 GB and 81.07 GB, respectively) with custom *.bed files and using the mafsInRegion utility (22). Subalignments were transformed into fasta format with the MAF to FASTA programme (version 1.0.1) in Galaxy (23). The taxa

present in these subalignments was then compared to those obtained with the BLAST-hit method.

We merged the chromosome 7 and 8 subalignments corresponding to the *polB* gene using the `–merge` function in MAFFT. The alignment was trimmed in trimAl version 1.4.rev22 (24) with the `–automated1` option (which selects the optimal trimming method for the alignment). We then selected the best nucleotide substitution model in ModelTest-NG version 0.1.7 (25) (TVM+Γ4 under the AIC, BIC and cAIC criteria). Having determined the orthology of the integrations, we conducted a phylogenetic analysis under a relaxed-molecular clock using divergence-time calibrations in BEAST 2 version 2.6.6 (26). For this, we used 13 calibration points obtained from TimeTree (27) (Table A1) using log-normal distributions, and then ran a Bayesian MCMC chain for 100,000,000 generations, sampling every 5,000th generation. We confirmed that the analysis had converged by inspecting the mixing and stationarity of posterior samples and ensuring that the Estimated Sample Sizes (ESSs) were greater than 200 for all parameters (burn-in of 25%).

Finally, we tested the possibility of selective constraints acting on the integrations. To this end, we calculated the pairwise genetic distances (measured as the observed proportion of nucleotide differences) between the taxa in the chromosome 7 and chromosome 8 subalignments separately, and compared these to 100 surrounding non-coding genomic regions of the same size in each case. We built two empirical distributions from these, and then tested the hypothesis that the distribution of *Maverick* genetic distances was distinct from the non-coding distribution. This was done with the a one-tailed, non-parametric

Kolmogorov-Smirnov test. An R script was developed for this purpose using functionality from the Ape package (28). We further checked if the human integrations fell within piRNA clusters reported for chromosomes 7 and 8 (29), but this resulted in no matches (Supplementary excel file).

Results

The *Maverick*-like integrations on chromosomes 8 and 7 are orthologous across the clade of placental mammals (Eutheria). The orthology of the regions was validated by the syntenic arrangement of the most proximal protein-coding genes, presence of the taxa in the whole-genome alignment of 120 mammals, and the relative arrangement of BLAST hits to different *Maverick* proteins which were consistent with the genetic organisation of vertebrate *Mavericks*. In addition, the taxonomic distributions resulting from the BLAST-hit and whole-genome alignment methods were consistent with each other (Supplementary excel file and multiple sequence alignments). However, some taxa were present in the whole-genome alignment corresponding to the *polB* marker which were not detected with the BLAST method, probably because of high divergence of the sequences (producing higher e-values than the cut-off). We could not detect orthologous integrations in the genomes of monotremes (*Ornithorhynchus anatinus*) or marsupials neither through BLAST nor in the whole-genome alignments.

Orthologues to the human chromosome 7 element were found in the genomes of primates, rodents, afrotherians (elephants, manatees and armadillo) and xenarthrans (sloths and

armadillos) (Figure 1). In addition, the *polB* marker was present in the whole genome-
alignment for scandentians (tree shrews), dermopterans (flying lemurs), lagomorphs
(hares) and two laurasiatherians: the star-nosed mole (*Condylura cristata*) and the
European hedgehog (*Erinaceus europaeus*). This finding is consistent with the presence of
hits to the minor capsid protein (but not the polymerase) in other laurasiatherians, namely
carnivores and cetaceans (Supplementary excel file). Remarkably, we detected an
additional hit to the integrase in the genome of the brown bear (*Ursus arctos*): as expected,
it occurs in the same reading frame as the minor capsid and does not show similarity to the
integrases of retroelements in Censor (30). Orthologues to the human chromosome 8
integration were found in the genomes of primates and xenarthrans (Linnaeus' two-toed
sloth, *Choloepus didactylus*). The *polB* marker again had a broader taxonomic distribution
in the whole-genome alignment and also included: scandentians, dermopterans,
lagomorphs and chiropterans (bats) (Figure 2). In this case, we could not detect orthologues
neither in rodents, afrotherians nor in other laurasiatherians.

All the integrations we analysed were highly degraded showing strong signs of being non-
functional at the protein level. Firstly, the open-reading frames are interrupted with
multiple early stop-codon and frameshift mutations. The elements also lack the terminal
inverted repeats characteristic of intact *Maverick* elements. In the case of the human
chromosome 8 orthologues, all of the region downstream of the *pz* gene seems to be absent.
In addition, the genetic distances of the elements are not significantly different from the
distances between non-coding regions in their surrounding genomic neighbourhood
(Kolmogorov-Smirnov test, $p > 0.99$, Figures A1 and A2), which suggests the elements

have been evolving neutrally. Interestingly, the chromosome 7/8 *Maverick* orthologues appear to have been lost in some mammals since they could not be detected by either method but were still present in related species. For example, the chromosome 7 integration appears to be absent from the genome of the Ugandan red colobus (*Piliocolobus tephrosceles*) while it is present in other colobine monkeys (*Trachypithecus francoisi*, *Rhinopithecus roxellana*). Similarly, the chromosome 8 integration seems to be absent from the genomes of the grey mouse lemur (*Microcebus murinus*) and Coquerel's sifaka (*Propithecus coquereli*), despite the *polB* marker being found in the northern greater galago (*Otolemur garnettii*) (this sequence is present in the whole genome alignment and shows a BLAST hit to the PolB of the common-box turtle with an e-value of $5e-4$).

The presence of the *polB* marker on both the chromosome 7 and 8 orthologues allowed us to estimate a joint phylogeny with the common ancestor of the elements represented by the root of the tree (Figure 3). The Bayesian estimate for the age of this common ancestor indicates it existed at the end of the Palaeozoic or start of the Mesozoic Eras (Figure 3, Table 1). Although we observed heterogeneous rates in the analysis, the inferred rates (mean = 0.0025, IQR = [0.0021, 0.0028]) agree with the overall pan-genomic substitution rates reported previously for mammals: 0.0027 nucleotide substitutions/site/million years (31).

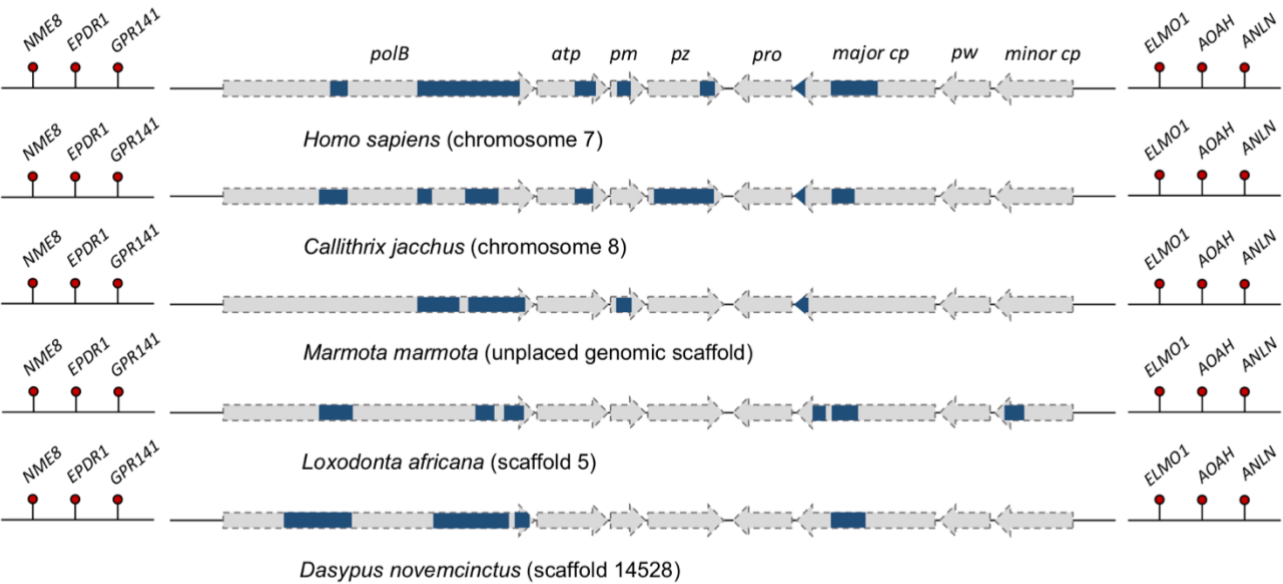


Figure 1. Comparison of the regions orthologous to the chromosome 7 *Maverick* integration in humans (*Homo sapiens*). tBLASTn hits (e-value < 0.05) to the proteins encoded by the common box-turtle (*Terrapene carolina*) *Maverick* are shown in blue. The arrow outlines represent the open reading frames of the box-turtle’s *Maverick*.

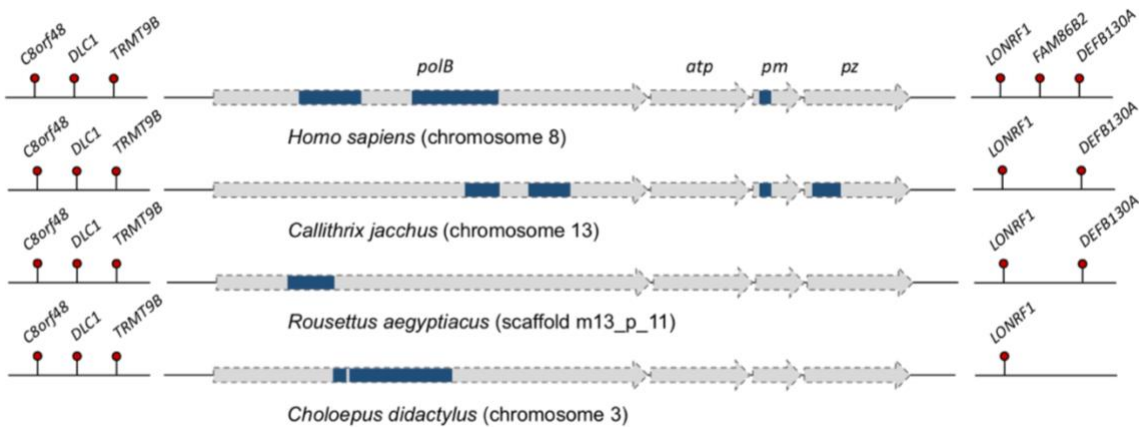


Figure 2. Comparison of the regions orthologous to the chromosome 8 *Maverick* integration in humans (*Homo sapiens*). tBLASTn hits (e-value < 0.05) to the proteins encoded by the Common box-turtle (*Terrapene carolina*) *Maverick* are shown in blue. The arrow outlines represent a segment of the open reading frames of the box-turtle's *Maverick*.

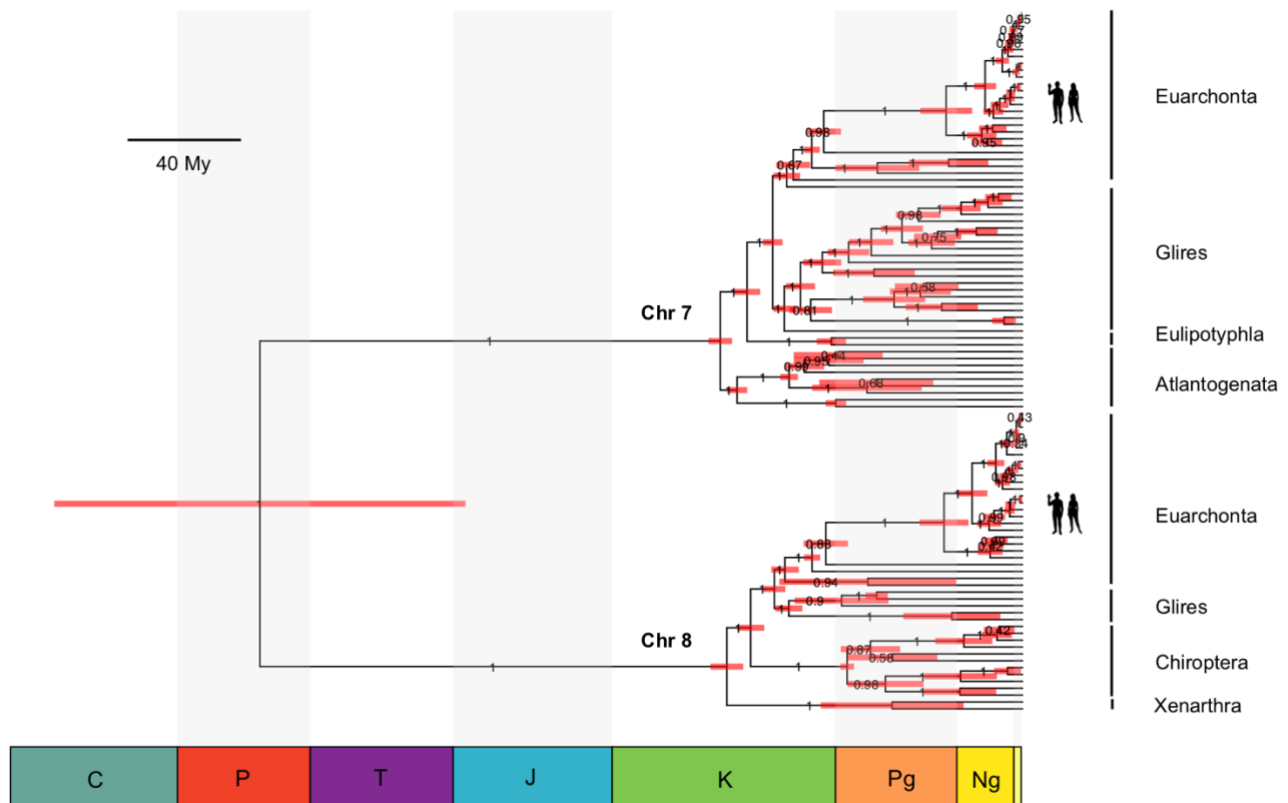


Figure 3. Phylogenetic tree of the mammalian polymerase B-like sequences orthologous to the insertions in human chromosomes 7 and 8. The horizontal node bars represent the 95% highest probability densities (HPD) for the age of the ancestor. Scale bar in million years (My).

Table 1. Posterior estimates for the age of the root, and the minimum ages of the chromosome 7/8 integrations (in million years from the present).

Node	Mean	Median	95% HPD
Root	268.61	264.05	[199.70–344.54]
Chromosome 7 CA	106.44	106.35	[102.44–110.78]
Chromosome 8 CA	104.12	104.06	[98.29–109.85]

Discussion

Our analyses have revealed that the two *Maverick*-like elements on chromosomes 8 and 7, with an estimated minimum age of 106 (102–111) and 104 (98–110) My, represent the oldest non-retroviral EVEs found in the human genome. An endogenous retrovirus-L element (ERV-L) of a comparable age (at least 104–110 My) has also been identified in humans and shown to be orthologous across the placental mammals (32). Comparison of the sequence divergence between the 5' and 3' LTRs of elephant and human (the 5' and 3' LTRs would be identical upon integration in the common ancestor), further suggests that the element may have integrated 24–36 My prior to the initial split between afrotherians and boreoeutherians (32). In terms of DNA viruses, an endogenous parvovirus-like element discovered in an intron of the *Ellis van Creveld syndrome 2* gene was shown to be present in primates, carnivores, ungulates and dolphins but not in afrotherians, giving it a minimum age of 98 My (33).

Discovery of the human element found on chromosome 8, together with the one found on chromosome 7, allowed us to gain a more detailed understanding of the evolutionary history of the elements before the diversification of placentals. Indeed, our time-calibrated phylogeny with both elements allowed us to infer the age of the root. This analysis suggests that *Mavericks* circulated in the ancestors of mammals 268 (199–344) Mya, which corresponds to the Carboniferous to Triassic periods. This is supported by the existence of *Maverick polb*-like sequences in the genomes of marsupials and monotremes albeit at non-orthologous positions with respect to the ones we have identified in humans (Table A2). Clear orthology of the chromosome 7 and 8 integrations across placental mammals together with their absence in marsupials and monotremes, suggest that the viruses integrated into the genome of the placental ancestor after the split with marsupials. Molecular estimates place this divergence around 172 (168–178) Mya (34), which is consistent with the age of the earliest fossil eutherian, *Juramaia sinensis*, from the Late Jurassic (160 Mya) of China (35, 36). Thus, it seems likely that active *Mavericks* persisted until the Jurassic Period in the genomes of the early eutherian ancestors.

It also seems that both chromosome 7 and 8 integrations had already been inactivated by the time of the most recent common ancestor of placentals (~105 Mya) (27), pointing to an older age of integration. This is evidenced by deletion of the genes downstream of *pz* in the chromosome 8 orthologues and the absence of the *pz*, *pro* (protease) and *pw* genes in the chromosome 7 insertions. In this respect, these *Maverick* insertions resemble those found in birds which are highly degraded and whose genes contain multiple inactivating mutations (16). We show that the two insertions found in humans and placental mammals

are likely to be non-functional since they do not seem to be under selection, they do not localise to piRNA clusters and have been lost altogether on several occasions in several species of mammals.

The reasons for the demise of *Mavericks* in mammals (and birds) remains an open question. Here we hypothesise several plausible scenarios. It has been suggested that *Mavericks* may function as a virophage-induced defence against the infection of large DNA viruses (37, 38), in particular iridoviruses (16). Iridoviruses are important pathogens of fish, amphibians and non-avian reptiles but they do not seem to infect either birds or mammals (39, 40). One possibility is that iridoviruses went extinct in mammals and birds as a result of this *Maverick* defence system. Once their viral hosts went extinct, endogenous virophages (which depend on a host virus for replication) would not have been able to further mobilise and this would have led to their degeneration. This hypothesis could be tested by performing iridovirus infection experiments on cell cultures of hosts which carry intact *Maverick* elements (teleost fish, amphibians or non-avian reptiles). Alternatively, endogenous *Mavericks* could have been co-opted as a defence against exogenous counterparts. Exaptation of endogenous viruses to serve this purpose has been reported extensively, for example in case of the *ev3*, *ev6* and *ev9* genes in chicken (41), *FvI* and *Fv4* genes in mice (42–45) and the Jaagsiekte endogenous retrovirus (enJS56A1) in sheep which are able to restrict exogenous viruses (46). Evolution of an effective antiviral defence system would lead to the extinction of the exogenous virus and degeneration of the defence locus once the selective pressure from the pathogen has been lifted (47). Therefore, the *Maverick* integrations found in humans and other placentals could represent

an antiviral defence system that has served its purpose and is now decaying. However, *Mavericks* have not been linked to any specific pathology so far and their exogenous stage has yet to be observed. Still another possibility is that these *Mavericks* were incidentally inactivated by host innate immune genes that evolved to fight other viral pathogens. For example, cytidine deaminases of the APOBEC3 family function as an antiviral defence in eutherian mammals and have been shown to target both DNA viruses and the DNA stage of retroviruses (48–50).

In addition to the viral arms-race scenarios, other hypotheses relating to changes in host biology could also explain the demise of *Mavericks* in the genomes of birds and mammals. It is possible that the receptor used for entry has been lost/acquired resistance mutations in these animals and the elements could no longer amplify in the germline by reinfection or cross-species transmissions. Support for this idea would require identification of the host cell receptor used for entry of *Mavericks* and assessment of its present state across vertebrates. Finally, the demise of *Mavericks* in these lineages might be linked to terrestrialsation and the origin of the amniotic egg, which may have limited the opportunities for the spread of these viruses and led to their ultimate degeneration. Although the natural history of *Mavericks* is still not understood, it seems they are most successful amongst aquatic organisms which suggests a water-borne mode of transmission (16, 51), perhaps they become active at the early stages when sperm and ova are released into the water by animals with external fertilisation. In fact, potentially active elements reach their highest diversity and copy numbers in fish (where horizontal transfers have also been detected), while only two intact elements have been discovered in amniotes (in turtles

and lizards) (16, 17). It therefore seems likely that the reduced opportunities for transmission could have led to a decrease in copy numbers which eventually made the elements prone to extinction.

Our results support a model where *Maverick* endogenous viruses were still circulating in the genomes of the stem mammals at the end of the Palaeozoic and even persisted in the genomes of the eutherian ancestors into the Jurassic Period. The elements found integrated on human chromosomes 8 and 7 represent the relics of ancient viruses that infected our ancestors >102 million years ago. Several hypotheses could explain the demise of *Mavericks* in mammals, which range from exaptation for an antiviral defence system followed by decay, inactivation by an innate immune mechanism, loss/mutation of the host cell receptor or changes to the biology of hosts which lowered their chances of transmission. Experimental studies focusing on the molecular virology of *Mavericks* and the virus-host interactions are needed to shed light on these issues.

Acknowledgements

This work was supported by the National Academy of Medicine of Venezuela and Pembroke College, Oxford (“Dr. José Gregorio Hernández” Award to J.G.N.B) and a European Research Council Award (101001623-PALVIREVOL to A. K.). We thank Michael Hiller for helping us retrieve the regions of the viral integrations from their whole-genome alignment of 120 mammals.

Data availability statement

The supplementary materials for this study have been deposited in Figshare (<https://dx.doi.org/10.6084/m9.figshare.17819708>). The R Code written to compare distributions of pair-wise genetic distances is available on GitHub (https://github.com/josegabrielnb/pair-wise_distributions).

References

1. Johnson WE. 2019. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nature Reviews Microbiology* <https://doi.org/10.1038/s41579-019-0189-2>.
2. Pritham EJ, Putliwala T, Feschotte C. 2007. Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* 390.
3. Kapitonov V v., Jurka J. 2006. Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America* 103.
4. Inoue Y, Saga T, Aikawa T, Kumagai M, Shimada A, Kawaguchi Y, Naruse K, Morishita S, Koga A, Takeda H. 2017. Complete fusion of a transposon and herpesvirus created the Teratorn mobile element in medaka fish. *Nature Communications* 8.
5. Inoue Y, Kumagai M, Zhang X, Saga T, Wang D, Koga A, Takeda H. 2018. Fusion of piggyBac-like transposons and herpesviruses occurs frequently in teleosts. *Zoological Letters* 4.
6. Aswad A, Katzourakis A. 2017. A novel viral lineage distantly related to herpesviruses discovered within fish genome sequence data. *Virus Evolution* 3.
7. Aimola G, Beythien G, Aswad A, Kaufer BB. 2020. Current understanding of human herpesvirus 6 (HHV-6) chromosomal integration. *Antiviral Research* <https://doi.org/10.1016/j.antiviral.2020.104720>.
8. Aswad A, Aimola G, Wight D, Roychoudhury P, Zimmermann C, Hill J, Lassner D, Xie H, Huang M-L, Parrish NF, Schultheiss H-P, Venturini C, Lager S, Smith GCS, Charnock-Jones DS, Breuer J, Greninger AL, Kaufer BB. 2020. Evolutionary

- History of Endogenous Human Herpesvirus 6 Reflects Human Migration out of Africa. *Molecular Biology and Evolution* <https://doi.org/10.1093/molbev/msaa190>.
9. Liu X, Kosugi S, Koide R, Kawamura Y, Ito J, Miura H, Matoba N, Matsuzaki M, Fujita M, Kamada AJ, Nakagawa H, Tamiya G, Matsuda K, Murakami Y, Kubo M, Aswad A, Sato K, Momozawa Y, Ohashi J, Terao C, Yoshikawa T, Parrish NF, Kamatani Y. 2020. Endogenization and excision of human herpesvirus 6 in human genomes. *PLoS Genetics* 16.
10. Henckaerts E, Dutheil N, Zeltner N, Kattman S, Kohlbrenner E, Ward P, Clément N, Rebollo P, Kennedy M, Keller GM, Linden RM. 2009. Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proceedings of the National Academy of Sciences of the United States of America* 106.
11. Daya S, Cortez N, Berns KI. 2009. Adeno-Associated Virus Site-Specific Integration Is Mediated by Proteins of the Nonhomologous End-Joining Pathway. *Journal of Virology* 83.
12. Krupovic M, Bamford DH, Koonin E v. 2014. Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. *Biology Direct* 9:1–7.
13. Krupović M, Bamford DH. 2008. Virus evolution: How far does the double β -barrel viral lineage extend? *Nature Reviews Microbiology* 6:941–948.
14. Krupovic M, Koonin E v. 2016. Self-synthesizing transposons: Unexpected key players in the evolution of viruses and defense systems. *Current Opinion in Microbiology* 31:25–33.
15. Krupovic M, Koonin E v. 2015. Polintons: A hotbed of eukaryotic virus, transposon and plasmid evolution. *Nature Reviews Microbiology* 13:105–115.
16. Barreat JGN, Katzourakis A. 2021. Phylogenomics of the Maverick Virus-Like Mobile Genetic Elements of Vertebrates . *Molecular Biology and Evolution* <https://doi.org/10.1093/molbev/msaa291>.
17. Starrett GJ, Tisza MJ, Welch NL, Belford AK, Peretti A, Pastrana D v, Buck CB. 2020. Adintoviruses: A Proposed Animal-Tropic Family of Midsize Eukaryotic Linear dsDNA (MELD) Viruses. *Virus Evolution* <https://doi.org/10.1093/ve/veaa055>.
18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410.
19. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Ridwan Amode M, Armean IM, Azov AG, Bennett R, Bhai J, Billis K, Boddu S, Charkhchi M, Cummins C, da Rin Fioretto L, Davidson C, Dodiya K, el Houdaigui B, Fatima R, Gall A, Giron CG, Grego T, Gujjarro-Clarke C, Haggerty L, Hemrom A, Hourlier T, Izuogu OG, Juettemann T, Kaikala V, Kay M, Lavidas I, Le T, Lemos D, Martinez JG, Marugán JC, Maurel T, McMahon AC, Mohanan S, Moore B, Muffato M, Oheh DN, Paraschas D, Parker A, Parton A, Prosovetskaia I, Sakthivel MP, Abdul Salam AI, Schmitt BM, Schuilenburg H, Sheppard D, Steed E, Szpak M, Szuba M, Taylor K, Thormann A, Threadgold G, Walts B, Winterbottom A, Chakiachvili M, Chaubal A, de Silva N, Flint B, Frankish A, Hunt SE, Iisley GR, Langridge N, Loveland JE, Martin FJ, Mudge JM, Morales J, Perry E, Ruffier M,

- 451 Tate J, Thybert D, Trevanion SJ, Cunningham F, Yates AD, Zerbino DR, Flicek P.
452 2021. Ensembl 2021. *Nucleic Acids Research* 49.
- 453 20. National Library of Medicine (US) NC for BI. 2021. Protein. Bethesda (MD).
- 454 21. Hecker N, Hiller M. 2020. A genome alignment of 120 mammals highlights
455 ultraconserved element variability and placenta-associated enhancers. *GigaScience*
456 9.
- 457 22. Kuhn RM, Haussler D, James Kent W. 2013. The UCSC genome browser and
458 associated tools. *Briefings in Bioinformatics* 14.
- 459 23. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Ech M, Chilton J, Clements
460 D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V,
461 Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The
462 Galaxy platform for accessible, reproducible and collaborative biomedical analyses:
463 2018 update. *Nucleic Acids Research* 46.
- 464 24. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for
465 automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*
466 25:1972–1973.
- 467 25. Darriba Di, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020.
468 ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein
469 Evolutionary Models. *Molecular Biology and Evolution* 37:291–294.
- 470 26. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M,
471 Gavryushkina A, Heled J, Jones G, Kühnert D, de Maio N, Matschiner M, Mendes
472 FK, Müller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D,
473 Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ. 2019.
474 BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis.
475 *PLoS Computational Biology* 15.
- 476 27. Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for
477 Timelines, Timetrees, and Divergence Times. *Molecular biology and evolution* 34.
- 478 28. Paradis E, Schliep K. 2019. Ape 5.0: An environment for modern phylogenetics and
479 evolutionary analyses in R. *Bioinformatics* 35.
- 480 29. Rosenkranz D, Zischler H, Gebert D. 2021. piRNAclusterDB 2.0: update and
481 expansion of the piRNA cluster database. *Nucleic Acids Research*
482 <https://doi.org/10.1093/nar/gkab622>.
- 483 30. Kohany O, Gentles AJ, Hankus L, Jurka J. 2006. Annotation, submission and
484 screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC*
485 *Bioinformatics* 7.
- 486 31. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ,
487 Meredith RW, Ödeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu
488 H, Yang W, Hu J, Xiao J, Yang Z, Liu Y, Xie Q, Yu H, Lian J, Wen P, Zhang F, Li
489 H, Zeng Y, Xiong Z, Liu S, Zhou L, Huang Z, An N, Wang J, Zheng Q, Xiong Y,
490 Wang G, Wang B, Wang J, Fan Y, da Fonseca RR, Alfaro-Núñez A, Schubert M,
491 Orlando L, Mourier T, Howard JT, Ganapathy G, Pfenning A, Whitney O, Rivas M
492 v., Hara E, Smith J, Farré M, Narayan J, Slavov G, Romanov MN, Borges R,
493 Machado JP, Khan I, Springer MS, Gatesy J, Hoffmann FG, Opazo JC, Håstad O,
494 Sawyer RH, Kim H, Kim KW, Kim HJ, Cho S, Li N, Huang Y, Bruford MW, Zhan
495 X, Dixon A, Bertelsen MF, Derryberry E, Warren W, Wilson RK, Li S, Ray DA,
496 Green RE, O'Brien SJ, Griffin D, Johnson WE, Haussler D, Ryder OA, Willerslev

- E, Graves GR, Alström P, Fjeldsø J, Mindell DP, Edwards S v., Braun EL, Rahbek C, Burt DW, Houde P, Zhang Y, Yang H, Wang J, Jarvis ED, Gilbert MTP, Wang J, Ye C, Liang S, Yan Z, Zepeda ML, Campos PF, Velazquez AMV, Samaniego JA, Avila-Arcos M, Martin MD, Barnett R, Ribeiro AM, Mello C v., Lovell P v., Almeida D, Maldonado E, Pereira J, Sunagar K, Philip S, Dominguez-Bello MG, Bunce M, Lambert D, Brumfield RT, Sheldon FH, Holmes EC, Gardner PP, Steeves TE, Stadler PF, Burge SW, Lyons E, Smith J, McCarthy F, Pitel F, Rhoads D, Froman DP. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346.
32. Lee A, Nolan A, Watson J, Tristem M. 2013. Identification of an ancient endogenous retrovirus, predating the divergence of the placental mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.
33. Liu H, Fu Y, Xie J, Cheng J, Ghabrial SA, Li G, Peng Y, Yi X, Jiang D. 2011. Widespread Endogenization of Densoviruses and Parvoviruses in Animal and Human Genomes. *Journal of Virology* 85.
34. dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings of the Royal Society B: Biological Sciences* 279.
35. Luo ZX, Yuan CX, Meng QJ, Ji Q. 2011. A Jurassic eutherian mammal and divergence of marsupials and placentals. *Nature* 476.
36. Bi S, Zheng X, Wang X, Cignetti NE, Yang S, Wible JR. 2018. An Early Cretaceous eutherian and the placental-marsupial dichotomy. *Nature* 558.
37. Katzourakis A, Aswad A. 2014. The origins of giant viruses, virophages and their relatives in host genomes. *BMC Biology* <https://doi.org/10.1186/s12915-014-0051-y>.
38. Koonin E v., Krupovic M. 2017. Polintons, virophages and transpovirons: a tangled web linking viruses, transposons and immunity. *Current Opinion in Virology* 25:7–15.
39. Williams T, Barbosa-Solomieu V, Chinchar VG. 2005. A Decade of Advances in Iridovirus Research. *Advances in Virus Research* [https://doi.org/10.1016/S0065-3527\(05\)65006-3](https://doi.org/10.1016/S0065-3527(05)65006-3).
40. Chinchar VG, Hick P, Ince IA, Jancovich JK, Marschang R, Qin Q, Subramaniam K, Waltzek TB, Whittington R, Williams T, Zhang QY. 2017. ICTV virus taxonomy profile: Iridoviridae. *Journal of General Virology* 98.
41. Robinson HL, Astrin SM, Senior AM, Salazar FH. 1981. Host Susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. *Journal of Virology* 40.
42. Jolicoeur P, Baltimore D. 1976. Effect of Fv-1 gene product on proviral DNA formation and integration in cells infected with murine leukemia viruses. *Proceedings of the National Academy of Sciences of the United States of America* 73.
43. Jolicoeur P, Rassart E. 1980. Effect of Fv-1 gene product on synthesis of linear and supercoiled viral DNA in cells infected with murine leukemia virus. *Journal of Virology* 33.

44. Kozak CA, Gromet NJ, Ikeda H, Buckler CE. 1984. A unique sequence related to the ecotropic murine leukemia virus is associated with the Fv-4 resistance gene. *Proceedings of the National Academy of Sciences of the United States of America* 81.
45. Ikeda H, Sugimura H. 1989. Fv-4 resistance gene: a truncated endogenous murine leukemia virus with ecotropic interference properties. *Journal of Virology* 63.
46. Murcia PR, Arnaud F, Palmarini M. 2007. The Transdominant Endogenous Retrovirus enJS56A1 Associates with and Blocks Intracellular Trafficking of Jaagsiekte Sheep Retrovirus Gag. *Journal of Virology* 81.
47. Aswad A, Katzourakis A. 2012. Paleovirology and virally derived immunity. *Trends in Ecology and Evolution* <https://doi.org/10.1016/j.tree.2012.07.007>.
48. Krishnan A, Iyer LM, Holland SJ, Boehm T, Aravind L. 2018. Diversification of AID/APOBEC-like deaminases in metazoa: Multiplicity of clades and widespread roles in immunity. *Proceedings of the National Academy of Sciences of the United States of America* 115.
49. Conticello SG, Thomas CJF, Petersen-Mahrt SK, Neuberger MS. 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. *Molecular Biology and Evolution* 22.
50. Ito J, Gifford RJ, Sato K. 2020. Retroviruses drive the rapid evolution of mammalian APOBEC3 genes. *Proceedings of the National Academy of Sciences of the United States of America* 117.
51. Haapa-Paananen S, Wahlberg N, Savilahti H. 2014. Phylogenetic analysis of Maverick/Polinton giant transposons across organisms. *Molecular Phylogenetics and Evolution* 78.