Involvement of non-LTR retrotransposons in the cancer incidence and lifespan of mammals

Marco Ricci, Valentina Peona[1], Cristian Taccioli[2]

[1] Department of Organismal Biology – Systematic Biology, Uppsala University, Uppsala, Sweden

[2] Department of Animal Medicine, Health and Production, University of Padova, Padova, Italy

Abstract

The natural occurrence of closely related species that show drastic differences in lifespan and cancer incidence raised the interest in finding the particular adaptations and genomic characteristics underlying the evolution of long lifespans. Studies on transposable elements (TEs) have more and more linked them to ageing and cancer development. In this study, we compared the TE content and dynamics in the genomes of four Rodent and six Chiroptera species that show very different lifespans and cancer susceptibility including the long-lived and refractory to cancer naked mole rat (*Heterocephalus glaber*), the long-lived fruit bats (*Pteropus vampyrus*, *Rousettus aegypticaus*) and the short-lived velvety free-tailed bat (*Molossus molossus*). By analysing the patterns of recent TE accumulation (TEs that are potentially currently active) in high-quality genome assemblies, we found that the shared genomic characteristics between long-lived species that are refractory to cancer, is the strong suppression, or negative selection against the accumulation, of non-LTR retrotransposons. All the short-lived species did show a recent accumulation of these TEs. Non-LTR retrotransposons have been often found to take part in the immune response of the host against

viral infections, but their dysregulation can lead to phenomena of “sterile inflammation” and “inflammageing”. Therefore, we hypothesise that the uncontrolled non-LTR retrotransposon activity is an important factor explaining the evolution of shorter lifespans in both Rodents and Chiroptera species and potentially in all mammals. Finally, these results suggest that non-LTR retrotransposons can be agents promoting cancer and ageing in mammals thus they may be targets of future oncological therapies.

Keywords: transposable elements, cancer, lifespan, ageing, transposons, retrotransposons, naked mole rat, bats, sterile inflammation.

Background

Transposable elements (TEs) are repeated mobile elements present in almost all eukaryotes [1, 2]. They can accumulate in genomes and account for most of the organism genome size, for example, the human genome is > 50% repetitive [3]. Many studies show that TEs are implicated in gene duplications, inversions, exon shuffling, gene expression regulation and may play an important role in the long-term evolution of organisms [4–12]. For example, the co-option of TE-related proteins gave rise to vertebrate acquired immune system and mammalian placenta [13–19]. Although some TE insertions have been co-opted by genomes, the overall TE activity can be disruptive especially at short timescale. Indeed, TEs can be responsible for several human diseases [7, 20, 21] like immunodeficiency [22], coagulation defects [23], cardiomyopathies [24], and muscular dystrophies [25, 26]. Most interestingly, the dysregulation of TEs in cancer cells may be also involved in the establishment, development and plasticity of cancer [27–32]. Furthermore, The dysregulated TE activity in cancer cells is so pervasive and accentuated that the methylation level of transposable elements is used as a biomarker for the malignancy of tumors [33–37]. Given the manifold effects of TEs on health,

it is reasonable to consider the TE activity a key factor able to influence lifespan differences between species [38].

To deepen the understanding of the relationship between TEs and longevity, we chose to perform a comparative genomic analysis of TE elements among mammalian species with different lifespans. TEs that are most likely to cause mutations and genomic instability are the ones currently active, i.e., able to move throughout the genome [38–41]. Since there are no transposition essays available for many organisms, we use the genetic divergence of the TE insertions from their consensus sequences as a proxy for their active/inactive state [12]. Among all the mammalian species for which genome assemblies are publicly available, we chose four species of Rodentia and six Chiroptera that matched the following criteria: 1) species showing a high-quality genome assembly to maximize the quality and quantity of transposable elements assembled [1]; 2) species belonging to the same taxonomical order to maximize their shared evolutionary history; 3) species showing similar body masses (**Figure 1**).

The experimental design presented here is basically the comparison of TEs activity in genomes of mammals with different lifespans. On firstly we choose a group of species that doesn't have physiological adaptation to defeat Peto's paradox. Bigger sized animals evolved mechanism contrasting peto's paradox human have the control of gene expression of telomerase [42]. The genome of African elephant underwent to an expansion in copy number of the onco-suppressor TP53 [43]. Small sized mammals don't face this issue providing models of longevity disentangled from Peto's paradox. We selected mammals with a body mass lower than 2 kg in order to investigate the relationship between lifespan and transposable elements avoiding bias related to high body mass adaptations.

It is generally observed that smaller mammals have a shorter lifespan with respect to larger mammals [44]. The most striking exceptions to this rule are found in the two orders of Rodentia and Chiroptera. In these taxa the presence of both long- and short-lived species with similar body sizes (**Figure 1**) provides the opportunity to investigate the relationship between TE activity, lifespan and cancer incidence while controlling for the body mass. Here we test the hypothesis that the most recent mutagenic activity of TEs promote cancer incidence and ageing related diseases.

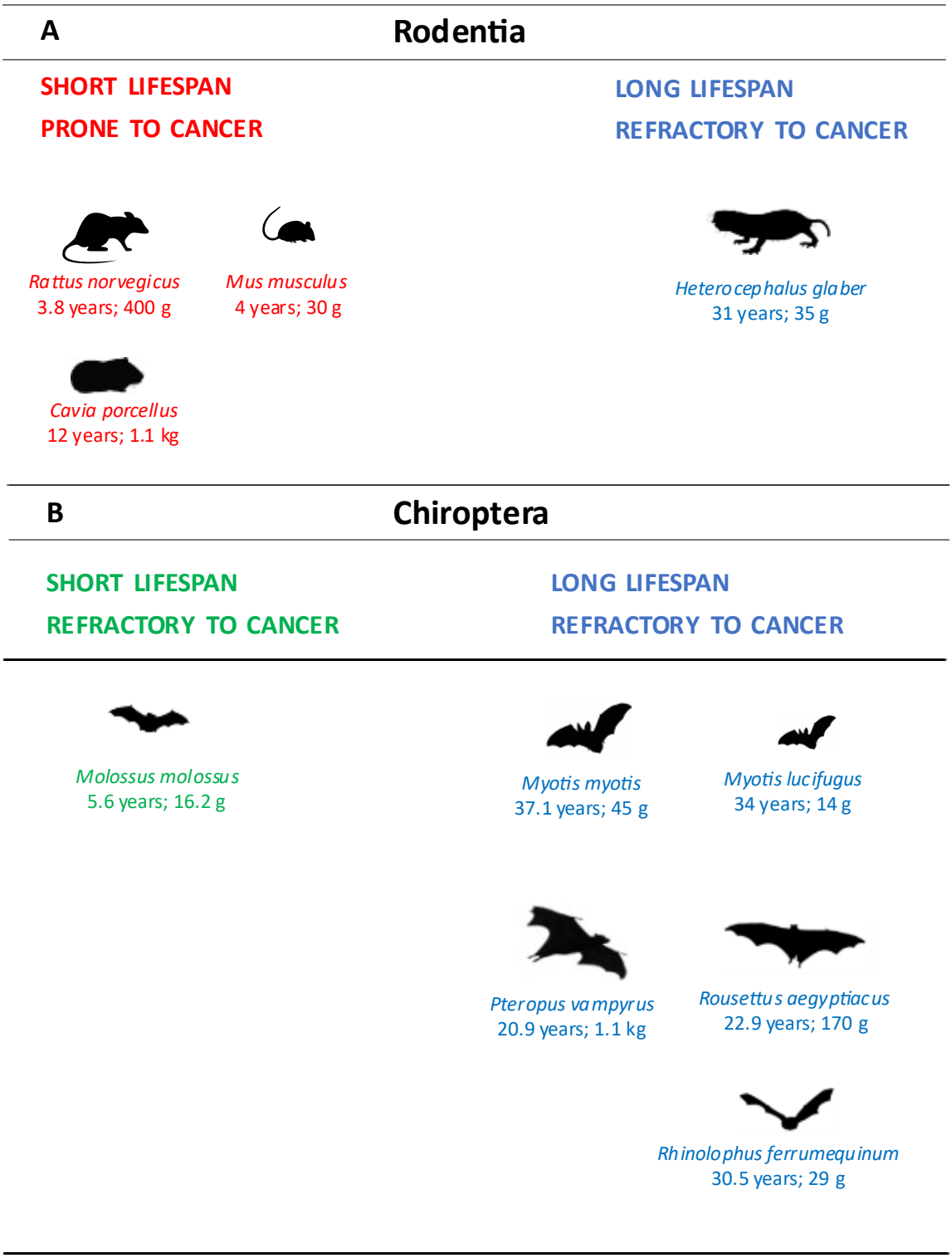

Figure 1. Species sampled and their maximum lifespan and body mass. A) Rodents with short lifespan and high cancer incidence are indicated in red while species with long lifespan and

low cancer incidence are indicated in blue. B) Bats with short lifespan and low cancer incidence are indicated in green while species with long lifespan and low cancer incidence are indicated in blue.

Results

Repeat content in short- and long-read based assemblies of naked mole rat

The repetitive content of genome assemblies is tightly linked to the sequencing technology adopted, in particular long reads substantially help to reconstruct complex repetitive regions that remain hidden using exclusively short reads [1]. For example, the naked mole rat (*Heterocephalus glaber*), there are two male genome assemblies available: the first version (HetGla_1.0 [45]) has been sequenced using the Illumina platform, whereas the most recent genome has been assembled combining PacBio reads together with the short ones (Heter_glaber.v1.7_hic_pac [46]). Therefore, we compared the two assembly versions to check the differences in the general repeat content to ultimately demonstrate the importance of perform TEs analysis with long reads-based genomes. First, we ran RepeatModeler2 on Heter_glaber.v1.7_hic_pac to produce a de-novo repeat characterization. The repeat library was used to mask both HetGla_1.0 and Heter_glaber.v1.7_hic_pac. In order to evaluate the improvement of TE detection in the last version of the assembly. This analysis shows that the total transposable element content is 34.27% in HetGla_1.0 and increases to 38.17% in Heter_glaber.v1.7_hic_pac (**Figure 2**). The latter percentage is more similar to the repeat content of the other rodents here analyzed: 37.2% in *Cavia porcellus*, 41.88 % *Mus musculus*, 41.13 % *Rattus norvegicus*. The TE category that presented the greatest increase in Heter_glaber.v1.7_hic_pac with respect to HetGla_1.0 was the LINE retrotransposons (from 20,02 % to 24,78 %). It is also possible to observe a slight reduction of the genomic portion annotated as Unknowns (-0.01%), DNA transposons (-0,21%), LTR retrotransposons (-0,3 %)

and SINEs (-0,34 %) in the PacBio assemblies. On the whole, the mobilome of the Naked mole rat showed a majority of retrotransposons and a minimal fraction of DNA transposons (less than 1/10 of the total TE content) in line with the repeat content of other rodents and mammals. Also, the respective abundance of retrotransposon categories corresponded with the other four rodents: a majority of LINEs is followed by LTR (mainly endogenous retroviruses) and SINE genomic elements.

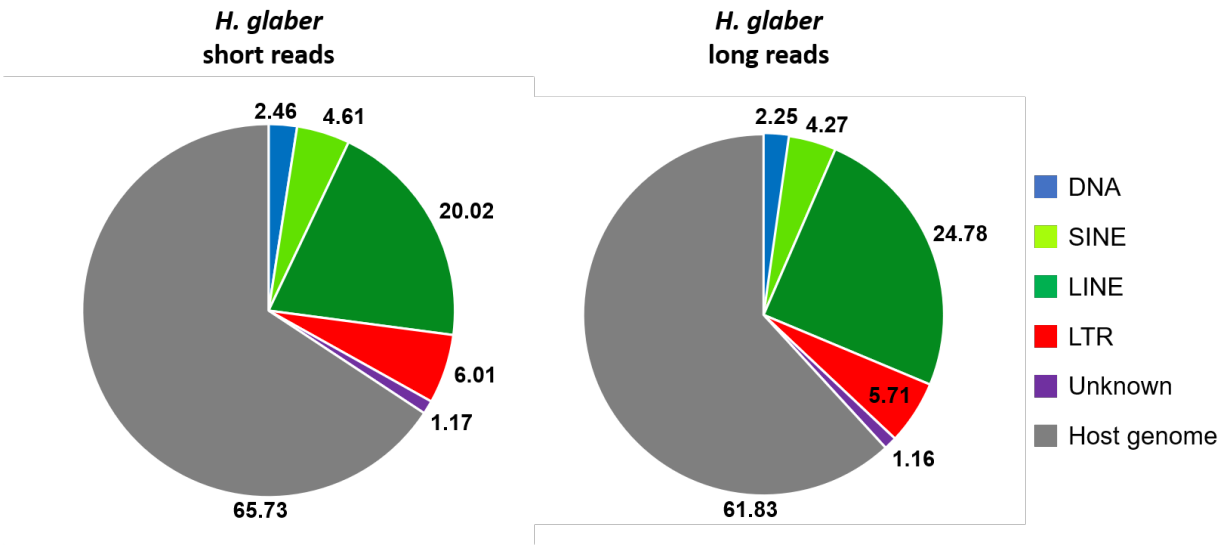

Figure 2 Repeat content comparison between the two genome assemblies of the naked mole rat *Heterocephalus glaber*. The pie charts show the percentage of the main transposable element categories. The portion of the host genome in grey comprises all the repetitive regions not annotated as transposable elements (e.g., tandem repeats and multi-copy gene families) as well as non-repetitive sequences. *H. glaber* short reads: HetGal_1.0, assembly size 2.6 Gb; *H. glaber* long reads: Heter_glaber.v1.7_hic_pac, assembly size 3 Gb.

Transposable element landscapes

To investigate how the repeat content of rodents and bats changed over time and detect the most recently active transposable elements, we used the RepeatMasker output file to generate a repeat landscape for each of the species (**Figure 3** and **Figure 4**). The repeat landscapes are a visualisation of the proportion of repeats (Y-axis) at different levels of divergence (X-axis) calculated as a Kimura 2-p distance between the insertions annotated and a consensus sequence. In rodent species, the repeat landscapes were mostly dominated by retrotransposons (**Figure 3**) and all of them showed an ancestral peak of activity (between 20 and 30% of divergence). *H. glaber* is the species with the biggest genome size and featured the biggest amount of ancient TEs (the peak overcame the 1.75 %, **Figure 3**). Part of those ancient TEs was made of relics of DNA transposons. *M. musculus* and *R. norvegicus* shared a peak of activity between 15 and 20% of divergence caused by an accumulation of unknown interspersed repeats. While *H. glaber* showed a second and more recent peak of activity between 10 to 5% of divergence dominated by LINEs, it also showed a dramatic accumulation drop from 5 to 0% of divergence which is specific to this species.

Table 1. For each rodent is presented the assembly size, the total content of transposable element and the specific amount of each TE category.

Species		***H. glaber***	***C. porcellus***	***M. musculus***	***R. norvegicus***
	assembly size Mb	3042	2723	2728	2648
Total TEs		38.16	36.01	41.86	41.76
DNA Transposons	**% bp**	2.3	1.7	1.1	1.1
	N° of elements	2,903,48	209910	143878	139359
SINE	**% bp**	4.3	5.5	7.6	7.2
	N° of elements	1115794	1308368	1495877	1448205

LINE	**% bp**	24.78	21.4	19.2	20.9
	N° of elements	1942418	715605	691600	654789
LTR	**% bp**	5.7	7.4	12.6	10.7
	N° of elements	519759	500519	866899	807643
Unknown	**% bp**	1.2	1.7	1.4	1.9
	N° of elements	120950	209910	44904	59085

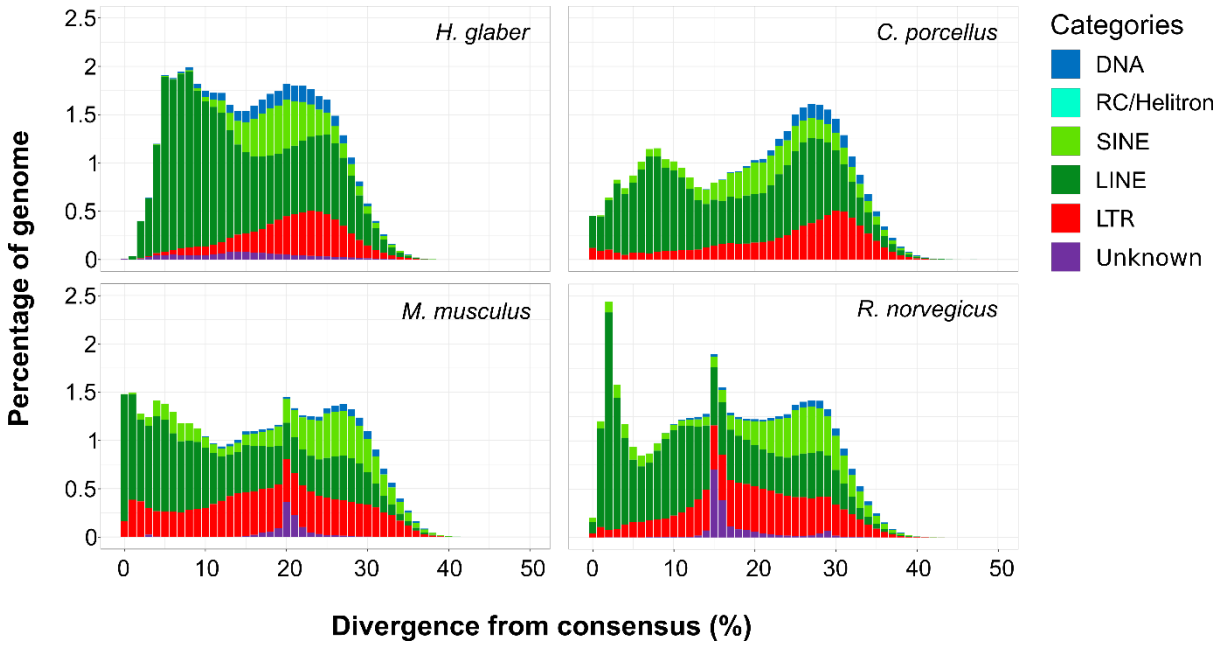

Figure 3. Repeat landscapes of rodent species. The Y-axis shows the fraction of genome occupied by transposable elements and the X-axis shows the genetic distance from the respective consensus sequences.

The TE landscapes of the six bat species analysed (**Figure 4**) showed a higher inter-specific diversity in comparison with the four rodents described above (**Figure 3**). The most peculiar feature of the bat landscapes was the presence of both retrotransposons (class I) and DNA transposons (class II). The two megabats belonging to the Pteropodidae family *Rousettus*

aegyptiacus and *Pteropus vampyrus* showed in general the lowest TE accumulation. *R. aegyptiacus, P. vampyrus* and *R. ferrumequinum* showed a shared high peak of activity that was located between the 25 and 27% of divergence from consensus. The other three species showed a quite different landscapes with the genus *Myotis* sharing a high activity peak at 15% of divergence that should correspond to the 8% peak in *M. molossus*. These three species belong to the clade of the Yangochiroptera (*M. myotis*, *M. lucifugus* and *M. molossus*) originated around 59 mya ago [47]. The two *Myotis* species show similar landscapes due to the closer phylogenetical relation. Although *M. lucifugus* had a greater activity and diversity of TE in its recent history (0 to 3% of divergence), *Myotis myotis* showed a longer accumulation of LTR retroposons in time with a peak between 4 and 5% of divergence from consensus. Among Chiroptera, the genus *Myotis* showed the most pronounced accumulation of class II TEs including both DNA transposase-dependent elements and Helitrons. In all the bat species here considered, the non-LTR retrotransposons (LINEs and SINEs) showed a "*glaber*-like" dynamic with a dramatic drop of activity between 0 and 3% of divergence from their consensus sequences.

Table 2. For each bat is presented the assembly size, the total content of transposable element and the specific amount of each TE category.

Species	***M. myotis***	***M. lucifugus***	***R. ferrumequinum***	***R. aegyptiacus***	***P. vampyrus***	***M. molossu***
Assembly size Mb	2003	2035	2075	1893	1996	2319
TE %						
Total TEs	32.33	31.66	34.29	29.08	28.84	40.6
DNA Transposons	4.96	5.12	5.32	3.78	3.81	3.2
HELITRONS	6.93	6.18	0.06	0.06	0.06	0.52
SINE	5.79	5.88	2.31	1.5	1.56	9,6
LINE	15.62	14.76	20.13	17.74	17.33	22,6

LTR	5.79	5.72	6.42	5.97	6.06	5,1
Unknown	0.17	0.19	0.1	0.08	0.08	0.1

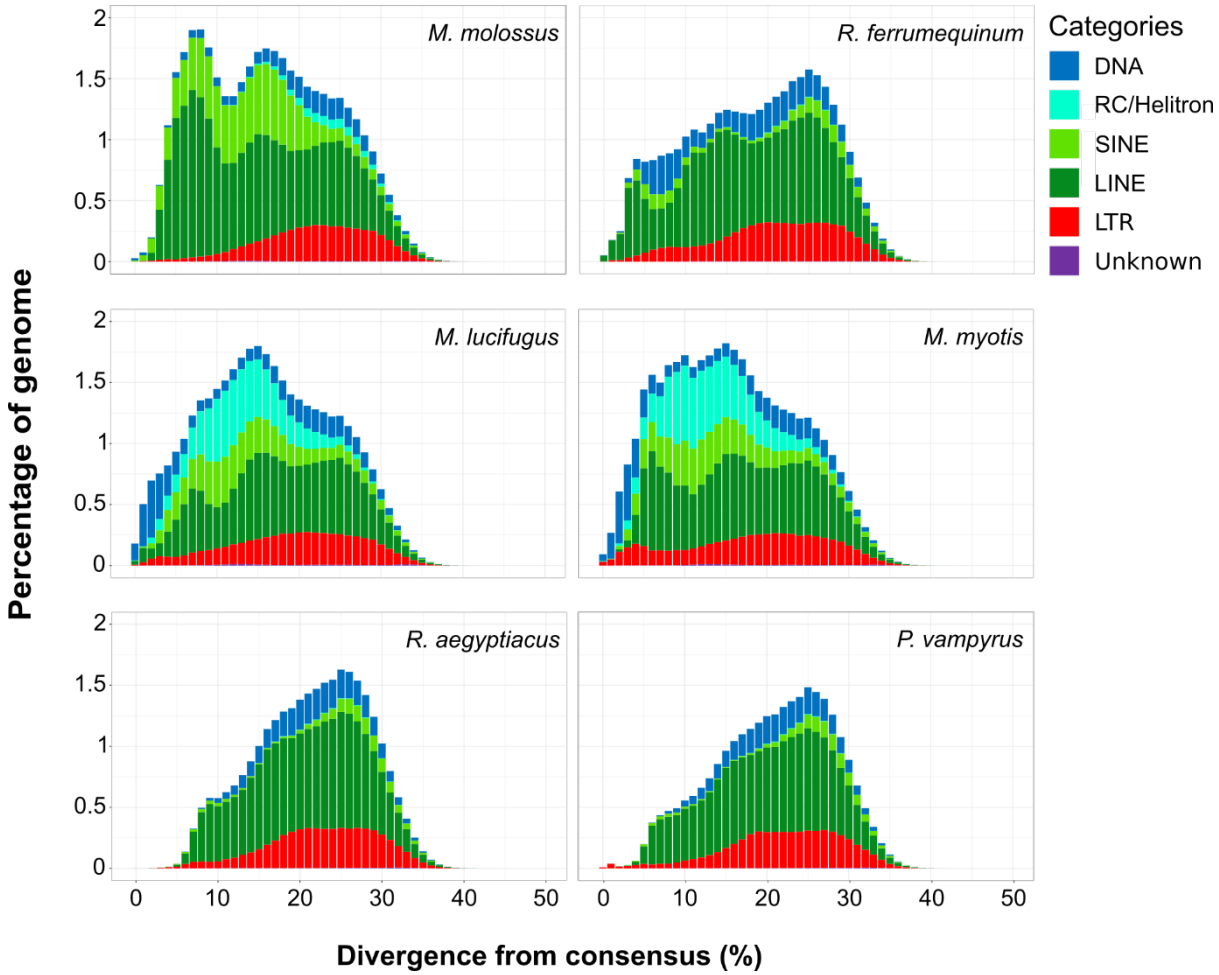

Figure 4. Repeat landscapes of bat species. The Y-axis shows the fraction of genome occupied by transposable elements and the X-axis shows the genetic distance from the respective consensus sequences.

We further explored the TE accumulation at low levels of divergence (more recent evolutionary times) by calculating the density of insertions (DI) from 0 to 3% of divergence. The density of insertion was estimated as the ratio between the number of non-LTR retrotransposon insertions

(LINEs and SINEs) found with a divergence ranging between 0 and 3% in a genome by RepeatMasker and the genome size itself. This measure allowed us to compare the magnitude of the TE accumulation and its potential impact between species with different genome sizes. When applying the density of insertion in rodents (**Figure 5**), we found that *H. glaber* showed the lowest DI (4,145). All the cancer-prone species, that also have a shorter lifespan than *H. glaber*, showed a higher DI. In general, the maximum lifespan of the four rodents involved in this study appeared to be inversely related to their DI of non-LTR retrotransposons. We calculated the density of insertions also for the six bat species considered (**Figure 6**). All the bat genomes showed a lower DI than rodents with the exception of *Molossus molossus* with a DI of 8,475. The lowest DIs of ~500 was observed in *R. aegyptiacus* and *P. vampyrus* (Megachiroptera). The remaining species *M. myotis, M. lucifugus* and *R. ferrumequinum* had a DI ranging between 2000 and 5000 retrelements per Gigabase . In general, the bats with a high lifespan (indicated in blue in **Figure 6**) presented a lower DI than *M. molossus.*

The LINE1 activity was furtherly investigated with the research of open reading frames (ORFs) containing a working reverse transcription machinery. All rodents, are equipped by complete ORFs with the exclusion of *H. glaber*. In bats, working Reverse transcriptase ORFs were found only in *M. myotis*, *M. lucifugus* and *R. ferrumequinum*.

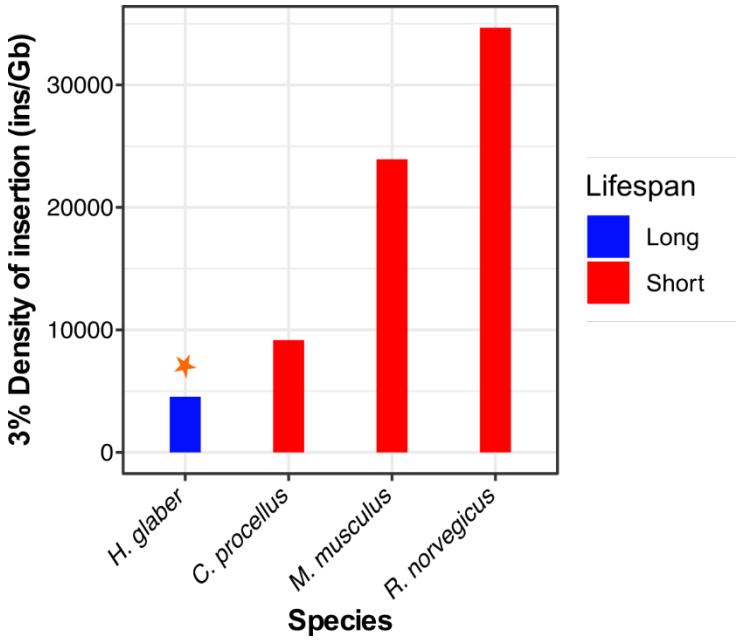

Figure 5. Barplots of the density of insertion in rodent species. The X axis shows the species analyzed in decreasing order of longevity. The Y-axis shows the number of recent insertions (non-LTR retrotransposons with a divergence from consensus lower than 3%) per gigabase in the analyzed genome. Blue bars indicate species with long lifespan and refractory to cancer. Red bars indicate species with short lifespan and prone to cancer. The orange star indicates the genomes in which intact LINE1 open reading frames were not found.

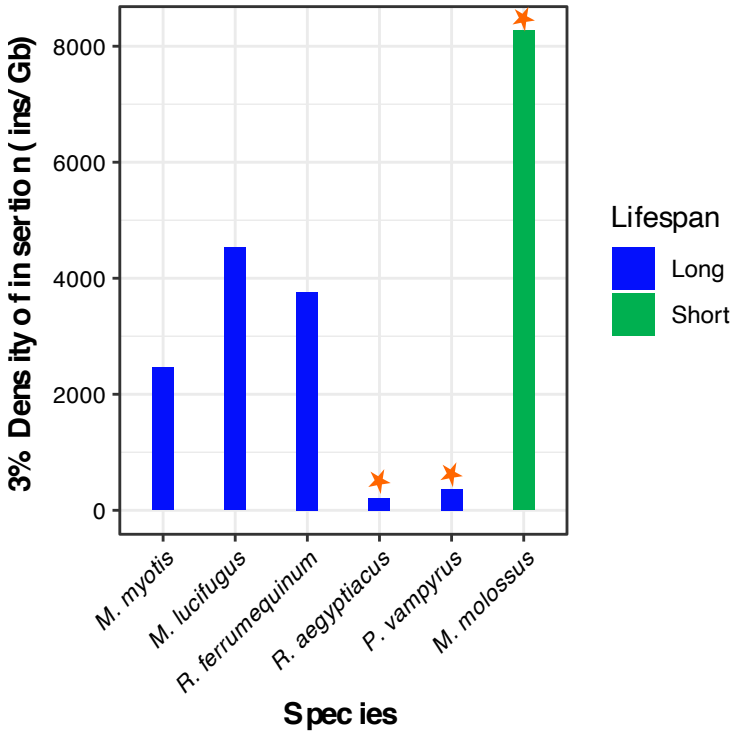

Figure 6. Barplots of the density of insertion in bat species. The X axis shows the species analyzed in decreasing order of longevity. The Y-axis shows the number of recent insertions (non-LTR retrotransposons with a divergence from consensus lower than 3%) per Gigabase in the analyzed genome. Blue bars indicate species with long lifespan and refractory to cancer. Red bars indicate species with short lifespan and prone to cancer. The orange star indicates the genomes in which intact LINE1 open reading frames were not found.

Discussion

Transposable element activity and accumulation can have manifold, deleterious or beneficial, effects on genomes and multiple studies have been linked transposable elements to ageing and the development of several diseases including cancer. [22, 30, 39, 48–50] In this study, we compared the transposable element dynamics of Rodents and Chiroptera to investigate whether their TE accumulation can explain the differences in cancer incidence and longevity observed in these species (**Figure 1**). The species considered in this study are known to have

different lifespans while sharing similar, small, body sizes. (> 2 kg). *H. glaber* is the rodent with the longest lifespan of 31 years and refractory to cancer while the other rodents show a lifespan of 12 (*C. porcellus*), 4 (*M. musculus*) and 3.8 (*R. norvegicus*) years. Chiroptera species, as far as it is currently known, are all refractory to cancer but show a large variability in lifespan (from 5. 6 years of *M. molossus* to 37 years of *Myotis myotis.* In the order of Chiroptera, cancer does not commonly occur as few confirmed cases have been reported [51, 52] and bat species generally show long lifespans for being small mammals [53]. Since TEs have been extensively linked to the development of cancer, we investigated the TE content of six bat species to see if it is possible to find shared features between long-lived bat species and long-lived rodent species.

To analyse the transposable element content of these species, we relied on the use of high-quality genome assemblies based on long reads that better represent the actual genomic repetitive content with respect to assemblies based on short reads [1]. Indeed, when comparing the two assembly versions (short and long reads) of *H. glaber*, the overall TE content increased from 25% [45] to 38% (**Figure 2**, **Table 1**) which is more similar to the total TE content generally found in other rodents (**Figure 3**, **Table 1**).

In Rodents, we identified *H. glaber* as the long-lived species refractory to cancer and *C. porcellus, M. musculus*, and *R. norvegicus* as short-lived species susceptible to cancer. We first compared their overall transposable element landscapes (**Figure 3**, **Table 1**). The short-lived species showed a continuous accumulation of TEs at any divergence level with high accumulation at very recent times (0-5% divergence). However, *H. glaber* genome showed a clear drop of accumulation of any TE in recent times. The longest living bat species here considered are *Myotis lucifugus* and *Myotis myotis* (**Figure 1**). Their two TE landscapes

showed the highest diversity of class II TEs of all bats with active elements of both DNA transposons and Helitrons in recent times (0-5% divergence) together with a recent accumulation of endogenous retroviruses. DNA transposons, Helitrons and ERVs seem to be active in *Myotis* genus but non-LTR retrotransposons (LINEs and SINEs) showed a drastic drop in accumulation at recent times (0-5% divergence) similarly to what can be observed in *H. glaber*. A similar drop of accumulation of non-LTR retrotransposons was also observed in the other four bat species.

As seen above (**Figure 3**, **Figure 5**, **Table 1**), the non-LTR retrotransposons are the most present types of TEs in Rodents and the most extensively investigated by biomedical research given that they are the only active TEs in the human genome [11]. Since the TEs that may pose the highest health threat to organisms are the ones presently active, we compared the most recent accumulation of non-LTR retrotransposons between long-lived and cancer-resistant species (*H. glaber*) and short-lived species commonly affected by cancer (**Figure 3**, **Figure 5**) using the density of insertion (DI). *H. glaber* showed the lowest DI of non-LTR retrotransposons followed by *C. porcellus* that has a lower cancer incidence than *M musculus* and *R. norvegicus* [54] (which showed a high DI) but higher than *H. glaber* (**Figure 5**). Given the pattern of DI observed in Rodents, the recent accumulation of non-LTR appears to correlate the lifespan of these species (**Figure 1**, **Figure 5**). On top of that, the genome of *H. glaber* does not contain intact ORFs of LINE retrotransposons which is another clue for the absence of non-LTR retrotransposon activity. The absence of intact LINE ORFs in the long read genome assembly, which should contain more insertions and maximize the portion of recent insertion, confirms the same observation previously done on a short read based genome assembly [55]. In bats, we observed that *M. molossus*, the species with the shortest lifespan, showed the highest DI, though no intact LINE ORFs were found in its genome. The other bat species considered

had a lower DI and no intact LINE ORFs were found in *R. aegyptiacus* and *P. vampyrus* which had the lowest DI among the considered chiropteran species. Previous works hypothesised that bats have a higher tolerance for the activity of transposable elements with alternative ways to dampen health or problematics to these species [41, 53]. Given our observation about the common drop of non-LTR retrotransposon accumulation in long-lived bats and *H. glaber*, we add to the previous hypothesis that it is the particular repression of the activity of non-LTR retrotransposon that confers resistance to cancer.

The case of *M. molossus*, being the bat species with the highest DI, shortest lifespan and no intact ORFs, is worth to discuss further. *M. molossus* genome has a high DI driven by a striking accumulation of non-autonomous SINEs that it is not observable in either chiropteran or rodent species. SINEs are retrotransposons that rely on active LINEs to complete their retrotransposition and since *M. molossus* genome does not present any intact LINE ORFs, it is reasonable to consider both LINES and SINEs inactive in this genome. The recently inserted SINEs and LINEs seem to be inactive, but they may still play a role in the physiology and ageing of *M. molossus*. A large number of very similar insertions across the chromosomes may cause structural polymorphisms (e.g., through non-allelic homologous recombination events) independently from the retrotransposition itself. For example, in the recent history of the human genome, LINE1-driven ectopic recombination events caused genomic rearrangements some of which responsible for diseases [14]. Moreover, it has been observed in the African killifishes that an accumulation of detrimental polymorphisms following a burst of TEs can undergo phases of relaxed selection were the mutations eventually promote age-associated diseases and reduced lifespan [56]. Among the killifishes, the turquoise killifish has a lifespan of 4-8 months (likely the shortest lifespan among vertebrates) but despite the accumulation of mutations that lead to a fast ageing, the species is fully adapted to his ecological niche demonstrating that a

rapid aging is not necessarily linked to a reduction in fitness. In addition to being considered DNA mutations, SINEs can also exert physiological effects by being transcribed. For example, in humans, SINE Alu elements have been found to be involved in the inflammatory pathway: upon viral infection, the interferon 1 mediates the overexpression of Alus which functions as an amplifier of the immune response by stimulating intracellular viral RNA sensor systems [57, 58]. Despite, the positive involvement of SINEs in the human immune response, the same upregulation of Alus, or the excessive sensitivity of the viral RNA sensors, has been linked to autoimmune diseases [59–61]. Furthermore, as non-coding RNAs are emerging as essential component of neuroinflammation, some studies have included Alu SINEs as possible promoters of neurodegenerative diseases [62]. More in general, the increased transcription of non-LTR retrotransposons in humans may contribute to the so-called "sterile inflammation", a phenomenon for which a chronic state of inflammation is triggered without the presence of any obvious pathogen and that is exacerbated with age ("inflammageing" [63]).

Many bat species share limited spaces in caves where they live very close to one another which make them particularly prone to viral infections. Not surprisingly, bats are the natural hosts of a vast number of viruses. Given the high number of viral infections that happen in bats, it is reasonable to think that bats have also evolved more than one strategy to counteract these infections. If bats evolved a system that involve the transcription of SINEs to amplify the immune response (similar to the interferon-Alu system in humans), then it is possible that in *M. molossus* the presence of SINEs has been positively selected but unintended side effects of this selection brought to the evolution of an extensive sterile inflammation and inflammageing that also caused the reduction of the species' lifespan. Consequently, it is possible that the excessive accumulation of SINEs exerted a selective pressure to, in turn, reduce the presence of active non-LTR retrotransposons and then be the cause of drastic accumulation drop in the

most recent times in the TE landscape of *M. molossus* genome (**Figure 5**). M. molossus may be an exceptional model for biogerontology.

Conclusions

An increasing number of studies are finding a link between TEs, ageing, and pathologies. As discussed in previous studies [39, 41] it is reasonable that genome instability derived by certain TEs could facilitate age-associated pathologies. The involvement of non-LTR retrotransposons in processes as the inflammation pathway explain the tight symbiosis between these mobile elements and the host genome, for this reason we think that in the long term they could accidentally act as promoting factor, more than as simple biomarker, in cancer evolution and other age-associated conditions.

Methods

Samples

In this study we analysed the high-quality genome assemblies of four Rodentia species and six Chiroptera species: *Mus musculus* GCA_000001635.8 [Genome Reference Consortium mouse reference 38], *Rattus norvegicus* GCA_000001895.4 and [64], *Cavia porcellus* GCA_000151735.1 https://www.broadinstitute.org, *Heterocephalus glaber* GCA_014060925.1 [46], *Rousettus aegyptiacus* GCA_014176215.1 [65] , *Rhinolophus ferrumequinum* GCA_014108255.1 [65], Molossus molossus GCA_014108415.1 [65] *Pteropus vampyrus* GCA_000151845.1 [66], *Myotis myotis* GCA_014108235.1 [65], *Myotis lucifugus* GCA_000147115.1 [66]. The species taken into consideration belong to the same taxonomical order and show a small body mass (in the range of 12g to 1.1 kg). The species were chosen from the same order to maximize their shared evolutionary history, physiological features and life history traits. The species were then further divided in long- and short-lived species (Figure 1). The genome assemblies analyzed are all based on Sanger or PacBio

technologies to maximize the assembled repetitive content. [67–70]. Naked mole rat High quality assembly was compared with the previous short reads based assembly GCA_000230445.1[45]. In order to estimate the differences in TE content.

Repetitive element libraries and annotation

All the species analysed have a species-specific repeat library already available but *Heterocephalus glaber*. Therefore, we made a *de-novo* repeat library of *H. glaber* using RepeatModeler2 with the option -LTRstruct [71] The resulting consensus sequences were merged with ancient TE subfamilies from the rodent specific library from RepBase https://www.girinst.org/repbase/

The genome assembly of *H. glaber* were masked with RepeatMasker (v. 4.1.0) and his customized library. *M. musculus, R. norvegicus* and *C. Porcellus* were masked using the rodents specific library. The Chiroptera genome assemblies were masked using the curated repeat library from Jebb et al. [65] with exception of *P. vampyrus* and *M. lucifugus* which annotations were directly downloaded from RepeatMasker datasets. The repeat annotations were then visualized using customized R script and the ggplot2.

Density of young retrotransposon insertions

To evaluate the density of recent retrotransposon insertions, we selected all the LINE and SINE hits annotated by RepeatMasker with a divergence from their consensus sequence less than 3%. For each species, the total number of the recent insertions were divided by the genome size (in gigabases) to obtain their density.

Intact ORF detection

To check the presence of intact LINE retrotransposons (therefore potentially active) we looked for complete ORF2 within the LINE annotated by RepeatMasker using the R script orf Cheker (https://github.com/jamesdgalbraith/OrthologueRegions/blob/master/orfChecker.R). The script looks for the presence of the reverse transcriptase and endonuclease domains in the same ORF to categorize it as intact.

Reference

1. Peona V, Weissensteiner MH, Suh A. How complete are "complete" genome assemblies?—an avian perspective. Mol Ecol Resour. 2018;18:1188–95.

2. Peona V, Blom MPK, Xu L, Burri R, Sullivan S, Bunikis I, et al. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. bioRxiv. 2019;:2019.12.19.882399. doi:10.1101/2019.12.19.882399.

3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921. doi:10.1038/35057062.

4. Hurst GDD, Schilthuizen M. Selfish genetic elements and speciation. Heredity (Edinb). 1998;80:2–8.

5. Ginzburg LR, Bingham PM, Yoo S. On the theory of speciation induced by transposable elements. Genetics. 1984;107:331–41.

6. Kazazian HH. Mobile Elements: Drivers of Genome Evolution. Science (80-). 2004;303:1626–32.

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

8. Jurka J, Bao W, Kojima KK. Families of transposable elements, population structure and the origin of species. Biol Direct. 2011;6:44. doi:10.1186/1745-6150-6-44.

9. Oliver KR, Greene WK. Mobile DNA and the TE-Thrust hypothesis: Supporting evidence from the primates. Mob DNA. 2011;2:8. doi:10.1186/1759-8753-2-8.

10. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18:71–86. doi:10.1038/nrg.2016.139.

11. Guichard E, Peona V, Malagoli Tagliazucchi G, Abitante L, Jagoda E, Musella M, et al. Impact of non-LTR retrotransposons in the differentiation and evolution of anatomically modern humans. Mob DNA. 2018;9:28. doi:10.1186/s13100-018-0133-4.

12. Ricci M, Peona V, Guichard E, Taccioli C, Boattini A. Transposable Elements Activity is Positively Related to Rate of Speciation in Mammals. J Mol Evol. 2018;86:303–10. doi:10.1007/s00239-018-9847-7.

13. Janoušek V, Laukaitis CM, Yanchukov A, Karn RC. The Role of Retrotransposons in Gene Family Expansions in the Human and Mouse Genomes. Genome Biol Evol. 2016;8:2632–50. doi:10.1093/gbe/evw192.

14. Lee J, Han K, Meyer TJ, Kim H-S, Batzer MA. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. PLoS One. 2008;3:e4047.

15. Moran J V, DeBerardinis RJ, Kazazian HHJ. Exon shuffling by L1 retrotransposition. Science. 1999;283:1530–4.

16. Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science. 2016;351:aac7247.

17. Kunarso G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010;42:631–4.

18. Koonin E V, Krupovic M. Evolution of adaptive immunity from transposable elements

combined with innate immune systems. Nat Rev Genet. 2015;16:184–92. doi:10.1038/nrg3859.

19. Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet. 2011;43:1154–9. doi:10.1038/ng.917.

20. Kazazian HHJ, Goodier JL. LINE drive. retrotransposition and genome instability. Cell. 2002;110:277–80.

21. Callinan PA, Batzer MA. Retrotransposable Elements and Human Disease. In: Genome Dynamics. 2006. p. 104–15. doi:10.1159/000092503.

22. Hancks DC, Kazazian Jr HH. Roles for retrotransposon insertions in human disease. Mob DNA. 2016;7:9. doi:10.1186/s13100-016-0065-9.

23. Beauchamp NJ, Makris M, Preston FE, Peake IR, Daly ME. Major structural defects in the antithrombin gene in four families with type I antithrombin deficiency--partial/complete deletions and rearrangement of the antithrombin gene. Thromb Haemost. 2000;83:715–21.

24. Yoshida K, Nakamura A, Yazaki M, Ikeda S, Takeda S. Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. Hum Mol Genet. 1998;7:1129–32.

25. Kondo-Iida E, Kobayashi K, Watanabe M, Sasaki J, Kumagai T, Koide H, et al. Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD). Hum Mol Genet. 1999;8:2303–9.

26. Narita N, Nishio H, Kitoh Y, Ishikawa Y, Ishikawa Y, Minami R, et al. Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. J Clin Invest. 1993;91:1862–7.

27. Carreira PE, Richardson SR, Faulkner GJ. L1 retrotransposons, cancer stem cells and

oncogenesis. FEBS J. 2014;281:63–73.

28. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res. 2014;24:1053–63.

29. Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, et al. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science (80-). 2014;345.

30. Burns KH. Transposable elements in cancer. Nat Publ Gr. 2017;17:415–24. doi:10.1038/nrc.2017.35.

31. Campbell PJ, Getz G, Korbel JO, Stuart JM, Jennings JL, Stein LD, et al. Pan-cancer analysis of whole genomes. Nature. 2020;578:82–93.

32. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet. 2020;52:306–19.

33. Ponomaryova AA, Rykova EY, Gervas PA, Cherdyntseva N V., Mamedov IZ, Azhikina TL. Aberrant Methylation of LINE-1 Transposable Elements: A Search for Cancer Biomarkers. Cells. 2020;9:1–19.

34. Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, et al. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of protooncogenes in human colorectal cancer metastasis. Gut. 2014;63:635–46.

35. Wolff EM, Byun HM, Han HF, Sharma S, Nichols PW, Siegmund KD, et al. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet. 2010;6.

36. Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-

small cell lung cancer. Int J Cancer. 2009;124:81–7.

37. Harada K, Baba Y, Ishimoto T, Chikamoto A, Kosumi K, Hayashi H, et al. LINE-1 Methylation Level and Patient Prognosis in a Database of 208 Hepatocellular Carcinomas. Ann Surg Oncol. 2015;22:1280–7.

38. Peona V, Palacios-Gimenez OM, Blommaert J, Liu J, Haryoko T, Jønsson KA, et al. The avian W chromosome is a refugium for endogenous retroviruses with likely effects on female-biased mutational load and genetic incompatibilities. bioRxiv. 2020.

39. Sturm Á, Ivics Z, Vellai T. The mechanism of ageing: Primary role of transposable elements in genome disintegration. Cell Mol Life Sci. 2015;72:1839–47. doi:10.1007/s00018-015-1896-0.

40. Chen H, Zheng X, Xiao D, Zheng Y. Age-associated de-repression of retrotransposons in the Drosophila fat body, its potential cause and consequence. Aging Cell. 2016;15:542–52.

41. Gorbunova V, Seluanov A, Mita P, McKerrow W, Fenyö D, Boeke JD, et al. The role of retrotransposable elements in ageing and age-associated diseases. Nature. 2021;596:43–53. doi:10.1038/s41586-021-03542-y.

42. Gorbunova V, Seluanov A. Coevolution of telomerase activity and body mass in mammals: From mice to beavers. Mech Ageing Dev. 2009;130:3–9.

43. Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, et al. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. Elife. 2016;5:1–30.

44. Speakman JR. Body size, energy metabolism and lifespan. J Exp Biol. 2005;208:1717–30.

45. Kim EB, Fang X, Fushan AA, Huang Z, Lobanov A V., Han L, et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature. 2011;479:223–7. doi:10.1038/nature10533.

46. Zhou X, Dou Q, Fan G, Zhang Q, Sanderford M, Kaya A, et al. Beaver and Naked Mole Rat Genomes Reveal Common Paths to Longevity. Cell Rep. 2020;32:107949. doi:10.1016/j.celrep.2020.107949.

47. Agnarsson I, Zambrana-Torrelio CM, Flores-Saldana NP, May-Collado LJ. A time-calibrated species-level phylogeny of bats (chiroptera, mammalia). PLoS Curr. 2011; FEB.

48. Morales-Hernández A, González-Rico FJ, Román AC, Rico-Leo E, Alvarez-Barrientos A, Sánchez L, et al. Alu retrotransposons promote differentiation of human carcinoma cells through the aryl hydrocarbon receptor. Nucleic Acids Res. 2016;44:4665–83.

49. Deininger PL, Batzer MA. Alu repeats and human disease. Mol Genet Metab. 1999;67:183–93.

50. Nguyen THM, Carreira PE, Sanchez-Luque FJ, Schauer SN, Fagg AC, Richardson SR, et al. L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution. Cell Rep. 2018;23:3730–40. doi:10.1016/j.celrep.2018.05.090.

51. Bradford C, Jennings R, Ramos-Vara J. Gastrointestinal leiomyosarcoma in an Egyptian fruit bat (Rousettus aegyptiacus). J Vet Diagnostic Investig. 2010;22:462–5.

52. Flanders AJ, Farina LL, Szivek A, Fox-Alvarez WA, Donnelly K, Hamel PES, et al. RENAL NEOPLASIA from A SINGLE POPULATION of PTEROPODID BATS. J Zoo Wildl Med. 2020;51:696–704.

53. Gorbunova V, Seluanov A, Kennedy BK. The World Goes Bats: Living Longer and Tolerating Viruses. Cell Metab. 2020;32:31–43. doi:10.1016/j.cmet.2020.06.013.

54. Albuquerque TAF, Drummond do Val L, Doherty A, de Magalhães JP. From humans to hydra: patterns of cancer across the tree of life. Biol Rev. 2018;93:1715–34.

55. Ivancevic AM, Kortschak RD, Bertozzi T, Adelson DL. LINEs between species: Evolutionary dynamics of LINE-1 retrotransposons across the eukaryotic tree of life. Genome Biol Evol. 2016;8:3301–22.

56. Cui R, Medeiros T, Willemsen D, Iasi LNM, Collier GE, Graef M, et al. Relaxed Selection Limits Lifespan by Increasing Mutation Load. Cell. 2019;178:385-399.e20.

57. Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M, et al. The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nat Immunol. 2004;5:730–7.

58. Hung T, Pratt GA, Sundararaman B, Towsend MJ, Chaivorapol C, Bhangale T, et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science (80-). 2015;350:455–9.

59. Volkman HE, Stetson DB. The enemy within: Endogenous retroelements and autoimmune disease. Nat Immunol. 2014;15:415–22.

60. Lagisquet J, Zuber K, Gramberg T. Retrotransposons Non-LTR. 2021;:1–19.

61. Ahmad S, Mu X, Yang F, Greenwald E, Park JW, Jacob E, et al. Breaching Self-Tolerance to Alu Duplex RNA Underlies MDA5-Mediated Inflammation. Cell. 2018;172:797-810.e13.

62. Polesskaya O, Kananykhina E, Roy-Engel AM, Nazarenko O, Kulemzina I, Baranova A, et al. The role of Alu-derived RNAs in Alzheimer's and other neurodegenerative conditions. Med Hypotheses. 2018;115 March:29–34. doi:10.1016/j.mehy.2018.03.008.

63. Franceschi C, Garagnani P, Parini P, Giuliani C, Santoro A. Inflammaging: a new immune–metabolic viewpoint for age-related diseases. Nat Rev Endocrinol. 2018;14:576–90. doi:10.1038/s41574-018-0059-4.

64. Gibbs RA, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 2004;428:493–521.

65. Science E, Kingdom U, Science B, Brook S, Brook S, Brook S, et al. ^Joint first authors *joint senior/corresponding authors emails. 2019;:1–34.

66. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. Nature. 2011;478:476–82.

67. Šklíba J, Lövy M, Koeppen SCW, Pleštilová L, Vitámvás M, Nevo E, et al. Activity of free-living subterranean blind mole rats Spalax galili (Rodentia: Spalacidae) in an area of supposed sympatric speciation. Biol J Linn Soc. 2016;118:280–91.

68. Hagan K. Heterocephalus glaber. Animal Diversity Web. Accessed October 27, 2020. 2013;:https://animaldiversity.org/accounts/Heterocephalu.

69. Armitage D. Rattus norvegicus. Animal Diversity Web. Accessed October 27, 2020. 2004;:https://animaldiversity.org/accounts/Rattus_norveg.

70. Ballenger L. Mus musculus. Animal Diversity Web. Accessed October 27, 2020. 1999;:https://animaldiversity.org/accounts/Mus_musculus/.

71. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020;117:9451–7. doi:10.1073/PNAS.1921046117.