Involvement of non-LTR retrotransposons in mammal cancer incidence and ageing

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 presence in nature of closely related species showing drastic differences in lifespan and cancer incidence has recently increased the interest of the scientific community on these topics. In particular, the adaptations and genomic characteristics underlying the evolution of cancer-resistant and long-lived species have recently focused on the presence of alterations in the number of non-coding RNAs, on epigenetic regulation and, finally, on the activity of transposable elements (TEs). In this study, we compared the content and dynamics of TE activity in the genomes of four rodent and six bat species exhibiting different lifespans and cancer susceptibility. Mouse, rat and guinea pig (short-lived and cancer-prone organisms) were compared with the naked mole rat (*Heterocephalus glaber*) which is the rodent with the longest lifespan. The long-lived and cancer-resistant bats of the genera *Myotis*, *Rhinolophus*, *Pteropus* and *Rousettus* were instead compared with the *Molossus*, which is instead a short-lived and cancer-resistant organism. Analyzing the patterns of recent accumulations of TEs in the genome in these species, we found a strong suppression or negative selection to accumulation, of non-LTR retrotransposons in long-lived and cancer-resistant organisms. On the other hand, all short-lived and cancer-prone species have shown recent accumulation of this class of TEs. Among bats, the *Molossus molossus* turned out to be a very particular species and, at the same time, an important model because, despite being susceptible to rapid ageing, it is resistant to

cancer. In particular, we found that its genome has the highest density of SINE (non-LTR retrotransposons), but, on the other hand, a total lack of active LINE retrotransposons. Our hypothesis is that the lack of LINEs presumably makes the *Molossus* cancer resistant due to lack of retrotransposition but, at the same time, the high presence of SINE, may be related to their short life span due to "sterile inflammation" and high mutation load. We suggest that research on ageing and cancer evolution should put particular attention to the involvement of non-LTR retrotransposons in these phenomena.

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

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 have shown 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 somatic cells may be also involved in the establishment, development and plasticity of cancer [27–32]. 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 several types of tumours [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]. The experimental design presented here is based on the comparison of the TE activity in genomes of mammals with different lifespans and cancer incidence. The TEs that cause mutations and genomic instability in the genomes are the ones currently active and able to move throughout the genome [38–41]. Since there are no transposition essays available for many organisms, we used the genetic divergence of the TE insertions (see **Methods**) from their consensus sequences as a proxy for their active or inactive state [12]. Among all the mammalian species for which genome assemblies are publicly available, we chose four species of Rodentia and six of Chiroptera that: 1) have a high-quality genome assembly (in order to maximize the quality and quantity of transposable elements assembled [1]); 2) belong to the same taxonomical order (to maximize their shared evolutionary history); and 3) have comparable body masses. In particular, the inclusion of body mass among the criteria of selection allowed us to work with species that have not evolved biological adaptations to counteract cancer incidence as a function of body size (Peto's paradox) [42]. In fact, animals with a big body mass evolved mechanisms to contrast the development of tumours such as the control of the telomerase expression [43] or the expansion in copy number of the onco-suppressor TP53 in *Loxodonta africana* (African elephant) [44]. Small mammals have a smaller number of cells and generally shorter life cycle [45], therefore they are disentangled from the Peto's paradox. For this reason, we selected mammals with a body mass lower than 2 kg to investigate the relationship between lifespan and transposable elements avoiding biases related to adaptations to large body masses. The species we used in our analyses represent both long and short lifespans (**Figure 1**) providing the opportunity to investigate the relationship between the TE activity and ageing. Here we test the hypothesis that the genomes of cancer-

prone and short-lived species present a higher load of recently inserted TEs than cancer-resistant and long-lived species.

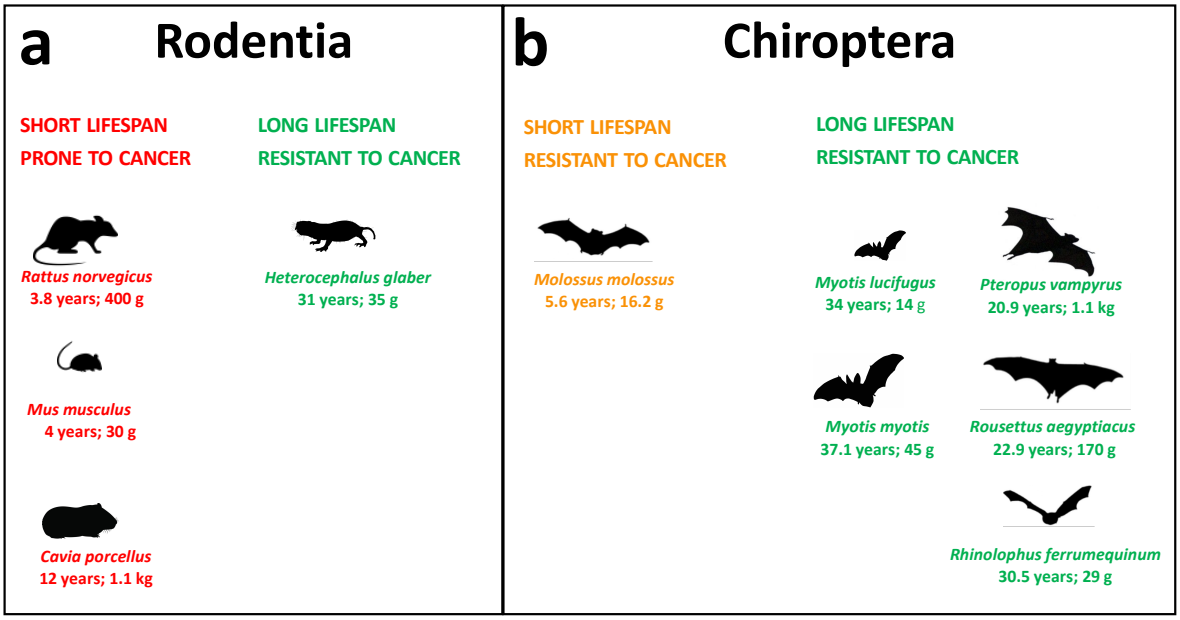

Figure 1. List of species analysed together with lifespan and body mass information. **a**) Rodents with short lifespan and high cancer incidence are indicated in red while species with long lifespan and low cancer incidence are indicated in green. **b**) Bats with short lifespan and low cancer incidence are indicated in orange while species with long lifespan and low cancer incidence are indicated in green. This information was retrieved from https://genomics.senescence.info/species/index.html and Speakman et al. [44].

Results

In this study we investigated the possible link between TEs, cancer incidence, and ageing by focusing on the accumulation patterns of TEs in genomes belonging to species showing different lifespans and cancer incidence (**Figure 1**). First, we collected genome assemblies of similar high quality and generated a TE library for the species that previously lacked one to both maximise the presence of TEs in the assemblies and their correct annotation. In particular, we showed that the last genome assembly version of *Heterocephalus glaber* (naked mole rat) based on long reads, yielded a higher percentage of the genome annotated as TEs with respect to its first version based on short reads (**Figure 2**, **Table 1**). Second, we proceeded with the comparison of the TE accumulation profiles between short- and long-lived rodents and

observed a reduced accumulation of non-LTR retrotransposons in the short-lived species (**Figure 3**). The same analysis in bats showed a high accumulation of class II transposons in all species (**Figure 4**, **Table 2**) but also highlighted a drop in the accumulation of non-LTR retrotransposons in the genomes of the long-lived bats (**Figure 4**). Finally, a detailed analysis of the most recent accumulation of non-LTR retrotransposons in rodents and bats showed a higher accumulation density in the species with the lowest lifespan (**Figure 5** and **6**).

TE 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 used. The long reads substantially help to reconstruct complex repetitive regions that remain hidden using exclusively short reads [1, 46]. For example, in the naked mole rat (*Heterocephalus glaber*), there are two male genome assemblies available: the first version (HetGla_1.0) sequenced using an Illumina platform and the latest version (Heter_glaber.v1.7_hic_pac) sequenced with PacBio. In the first analysis performed on HetGla_1.0 [47], the total TE content was 25% of the genome but such estimate is not available for Heter_glaber.v1.7_hic_pac [48]. Therefore, we produced a *de novo* repeat library based on the last assembly version and then annotated it. First, we ran RepeatModeler2 on Heter_glaber.v1.7_hic_pac to produce the *de novo* repeat library.

Then we used the newly produced repeat library to annotate the transposable elements in both assembly versions with RepeatMasker. The RepeatMasker annotation showed that the total transposable element content was 34.27% in HetGla_1.0 that increased to 38.17% in Heter_glaber.v1.7_hic_pac (**Figure 2**).

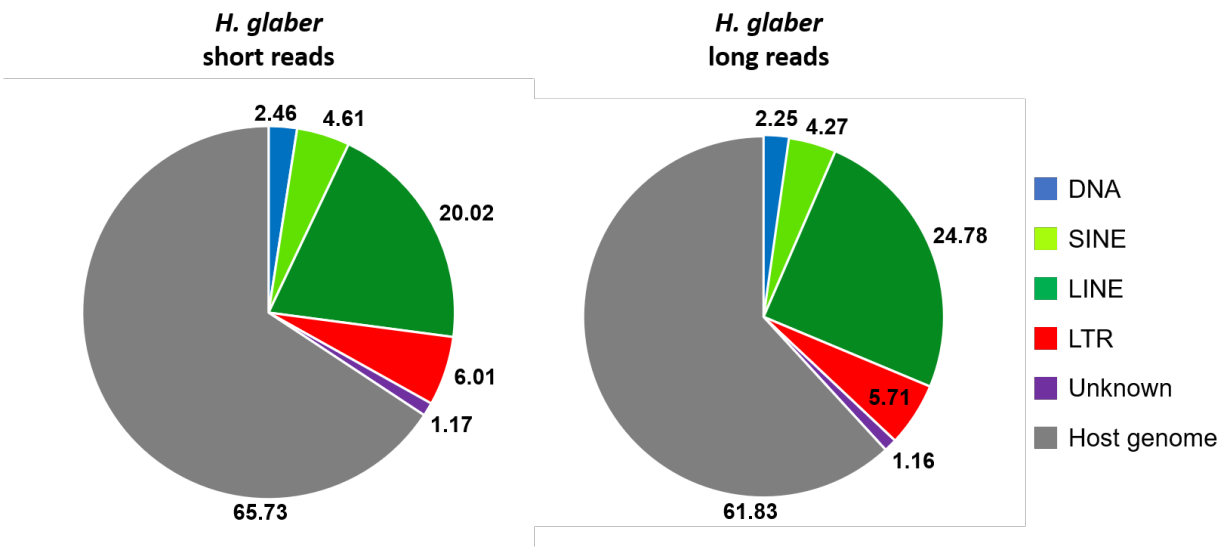

Figure 2. Transposable element content comparison between the two genome assemblies of the naked mole rat. The pie charts show the percentage of the main transposable element categories. The portion of the genome in grey comprises all the repetitive regions not annotated as transposable elements (e.g., 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.

This latter percentage (38.17%) is more similar to the TEs content of the other rodents here analysed: 37.2% in *Cavia porcellus*, 41.88% in *Mus musculus*, and 41.13% in *Rattus norvegicus*. The TEs that showed the greatest increase between the two assemblies were the LINE retrotransposon (from 20.02% to 24.78%). In Heter_glaber.v1.7_hic_pac was also possible to observe a slight reduction of TEs classified as Unknowns (-0.01%), DNA transposons (-0.21%), LTR retrotransposons (-0.3 %) and SINEs (-0.34 %). On the whole, the mobilome of the naked mole rat showed a majority of retrotransposons and a minimal fraction of DNA transposons (less than 10% of the total TE content) in line with the TE content of other rodents and mammals.

Transposable element analysis in rodents and bats

To investigate the differences in TE content between the rodent and bat species, we ran RepeatMasker on all their genome assemblies and the summary of the TE annotations are shown in **Table 1** and **Table 2**.

Table 1. The assembly size and the percentage of genome annotated as transposable elements are shown for each rodent genome.

Species	*H. glaber*	*C. porcellus*	*M. musculus*	*R. norvegicus*
Assembly size (Gb)	3.042	2.723	2.728	2.648
Total TE content (%)	38.16	36.01	41.86	41.76
DNA transposons (%)	2.3	1.7	1.1	1.1
SINE (%)	4.3	5.5	7.6	7.2
LINE (%)	24.78	21.4	19.2	20.9
LTR (%)	5.7	7.4	12.6	10.7
Unknown (%)	1.2	1.7	1.4	1.9

Table 2. The assembly size and the percentage of genome annotated as transposable elements are shown for each bat genome.

Species	*M. myotis*	*M. lucifugus*	*R. ferrumequinum*	*R. aegyptiacus*	*P. vampyrus*	*M. molossus*
Assembly size (Gb)	2.003	2.035	2.075	1.893	1.996	2.319
Total TE content (%)	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

The genomes of the rodents and bats analysed here were mainly composed by class I retrotransposons (SINEs, LINEs, LTRs). The main difference between the two orders was that all six bats had a higher percentage of class II transposons (DNA transposons and Helitrons) than rodents, with the minimum percentage present in *M. molossus* (3.72%) and the maximum percentage in *M. myotis* and *M. lucifugus* (>11%; **Table 2**). Rodents, in general, showed a limited amount of class II transposons with the maximum abundance present in the naked mole

rat (2.3 %). While the naked mole rat genome was the largest, it showed the lowest percentage of SINEs and LTR retrotransposons, and the highest percentage of LINEs among the analysed rodents (**Table 1**). Notably, *Myotis myotis* and *Myotis lucifugus* revealed an exceptional accumulation of Helitrons (6.93 and 6.18 % respectively) in comparison with the other bats that showed an abundance of Helitrons below 1%. The short-lived *M. molossus* showed contemporarily the largest genome size and the highest TE content (**Table 2**), in particular, the highest percentage of non-LTR retrotransposons (SINEs and LINEs).

To investigate how the TE content of rodents and bats changed over time and detect the most recently active transposable elements, we used the RepeatMasker annotation (see **Methods**) to generate a TE landscape for each of the species. The TE landscapes are a visualisation of the proportion of repeats in base pairs (Y-axis) at different levels of divergence (X-axis) calculated as a Kimura 2-p distance [49] between the insertions annotated and their respective consensus sequences (**Figure 3** and **Figure 4**). The TE landscapes in rodents were mostly dominated by retrotransposons and all of them showed an ancestral peak of accumulation between 20 and 30% of divergence (**Figure 3**). A small percentage of those ancient TEs was composed of relics of DNA transposons (blue in **Figure 3**). Notably, the cancer-resistant rodent naked mole rat showed the highest accumulation of retrotransposition dominated by LINEs (green) between 5 and 10% of divergence followed by a dramatic drop corresponding to the most recent history of this genome (**Figure 3**).

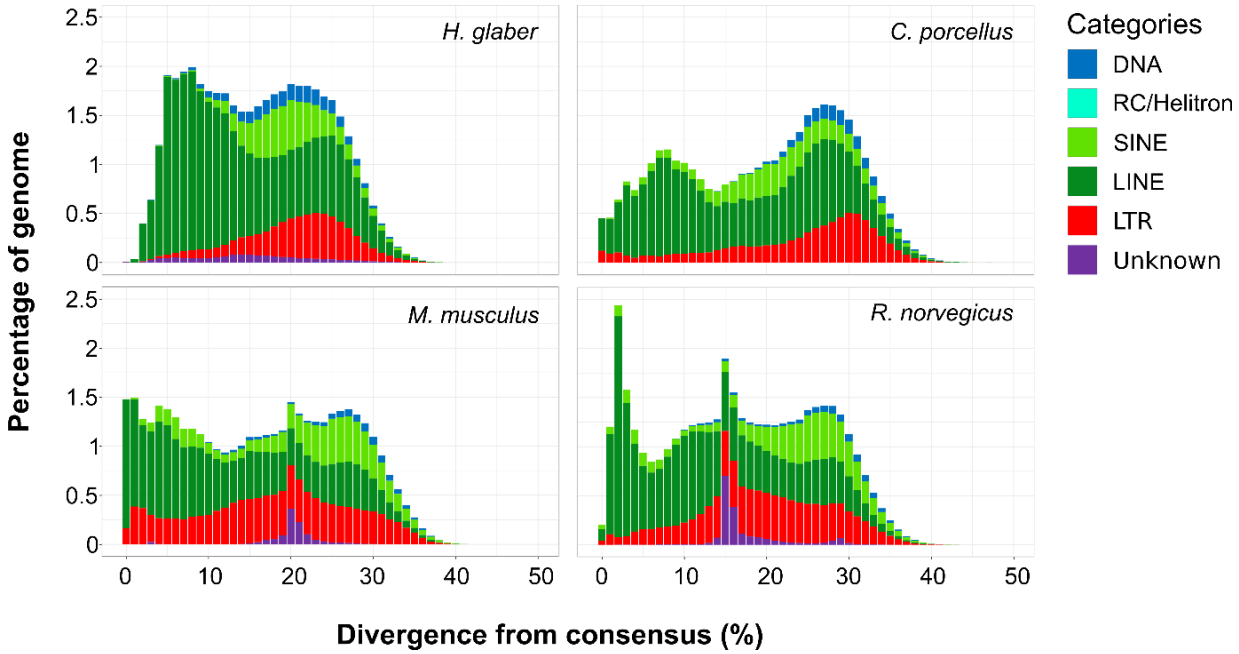

Figure 3. Transposable element landscapes of rodent species. The X-axis shows the genetic distance between the transposable element insertions and their consensus sequences, and the Y-axis shows the percentage of the genome occupied by transposable elements.

The three cancer-prone rodents showed an ongoing accumulation of LTR retrotransposons (red) and LINEs. *M. musculus* and *R. norvegicus* shared a peak of accumulation between 15 and 20% of divergence made of unknown interspersed repeats (purple). *R. norvegicus* was the only rodent that maintained a stable rate of retrotransposition of SINEs (light green).

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 difference of the bat landscapes was the higher amount of class II transposons (**Table 2**). The two megabats belonging to the Pteropodidae family (*Rousettus aegyptiacus* and *Pteropus vampyrus*) presented the lowest TE accumulation (**Table 2**). Notably, *R. aegyptiacus, P. vampyrus* and *R. ferrumequinum* showed a shared peak of accumulation at 25% of divergence from consensus. The other three species (*M. myotis*, *M. lucifugus* and *M. molossus*) belong to the clade of Yangochiroptera evolved around 59 mya ago [50] and showed different landscapes. The two species belonging the *Myotis* genus shared their highest peak of

accumulation at 15% of divergence, whereas *Molossus* had a second more recent species-specific peak at 8% of divergence. The two *Myotis* species showed similar landscapes likely due to the closer phylogenetical relationship. While *M. lucifugus* had a higher accumulation and diversity of TEs in its recent history (0-3% divergence), *Myotis myotis* had a longer history of accumulation of LTRs with a peak between 4 and 5% of divergence from consensus. In Chiroptera, the genus *Myotis* had the most pronounced accumulation of DNA transposons and Helitrons (light blue). In all the bats here considered, the non-LTR retrotransposons (LINEs and SINEs) showed a "*glaber*-like" dynamic with a decrease of retrotransposition in correspondence of their most recent evolutionary history (**Figure 4**).

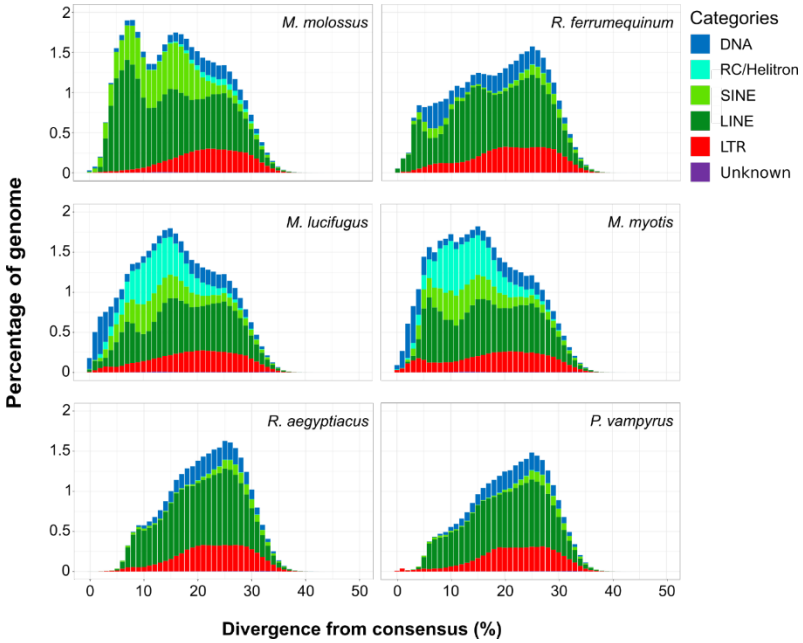

Figure 4. Transposable element landscapes of rodent species. The X-axis shows the genetic distance between the transposable element insertions and their consensus sequences, and the Y-axis shows the percentage of the genome occupied by transposable elements.

Activity of non-LTR retrotransposons in rodents and bats

We further explored the non-LTR retrotransposon accumulation (see **Methods**) 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 (SINEs and LINEs) insertions (**Figure 5** and **6**) and the genome size in Gigabases. The density of insertion is a

sensitive parameter to compare the magnitude of the recent TE accumulation and its potential impact between species with different genome sizes [12]. When applying this measure in rodents (**Figure 5**), we found that the naked mole rat showed the lowest DI (4,145). All the cancer-prone species, that also have a shorter lifespan than the naked mole rat, showed a higher DI (**Figure 5**).

Then we calculated the density of insertions for the six bat species considered (**Figure 6**). All the bat genomes showed a lower DI than rodents with the exception of *Molossus molossus* that presented a DI of 8,475. The lowest DI (<1,000) was observed in *R. aegyptiacus* and *P. vampyrus*. The remaining species *M. myotis, M. lucifugus* and *R. ferrumequinum* had a DI ranging between 2000 and 5000. In general, the long-lived bats (indicated in blue in **Figure 6**) presented a lower DI compared to the short-lived *M. molossus* (**Figure 6**).

To investigate the possible presence of not only recently accumulated but active autonomous non-LTR retrotransposons we looked for the presence of LINE-related open reading frames (ORFs) with intact protein domains. By doing this, we found that all the analysed rodents genomes present intact LINE ORFs with the exclusion of the naked mole rat (**Figure 5**). In bats, intact LINE ORFs were found only in *M. myotis*, *M. lucifugus* and *R. ferrumequinum* (**Figure 6**).

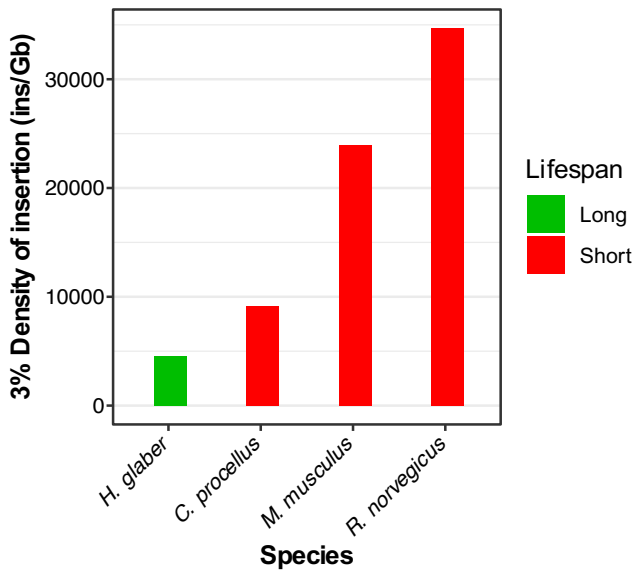

Figure 5. Density of insertion of recent non-LTR retrotransposons in rodents. The X-axis shows the species analyzed in decreasing order of longevity. The Y-axis shows the density of recent insertions per Gigabase (<3% divergence from consensus). Green indicates cancer-resistant and long-lived species. Red indicates cancer-prone and short-lived species. The black skull indicates the genomes in which intact LINE open reading frames were not found.

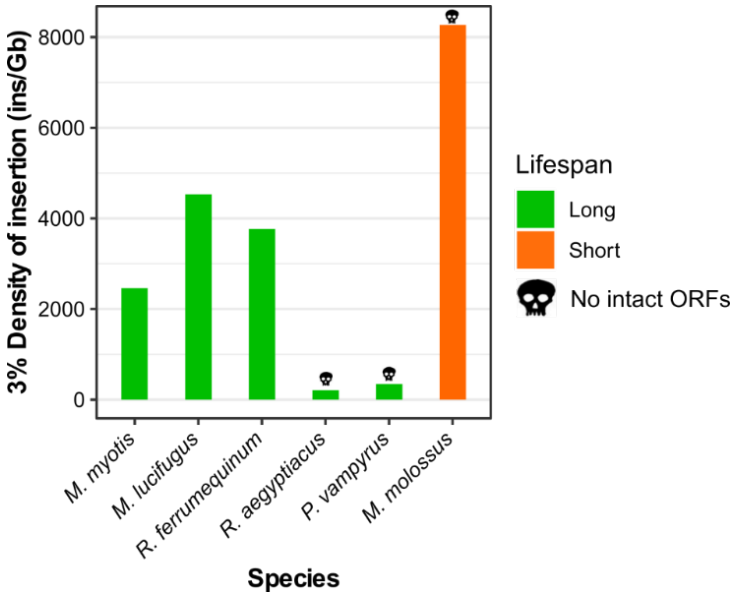

Figure 6. Density of insertion of recent non-LTR retrotransposons in bats. The X-axis shows the species analyzed in decreasing order of longevity. The Y-axis shows the density of recent insertions per Gigabase (<3% divergence from consensus). Green indicates cancer-resistant and long-lived species. Orange indicates cancer-resistant and short-lived species. The black skull indicates the genomes in which intact LINE open reading frames were not found.

Discussion

Transposable element activity and accumulation can have manifold effects on genomes. Multiple studies have been linked transposable elements to ageing and the development of several diseases including cancer [22, 30, 31, 39, 51–54]. Here, we have studied the relationship between transposable elements and two different aspects of mammal life: longevity and cancer incidence. In rodents the short lifespan is associated with the presence of cancer [55]. On the other hand, bats are considered cancer-resistant species [56, 57] (**Figure 1**). The animals 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 known (31 years) and resistant to cancer while the other rodents show shorter lifespans of 12 (*C. porcellus*), 4 (*M. musculus*) and 3.8 (*R. norvegicus*) years. Chiroptera species, as far as it is currently known, are resistant to cancer (only a few confirmed cases have been reported [58, 59]) and show long lifespans for being small mammals [56] with a large variability (from 5.6 years of *M. molossus* to 37 years of *Myotis myotis*). 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 and 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]. The use of long read-based genome assemblies is particularly important when analysing young TE insertions (as in this study) that are highly homogeneous and tend to be underrepresented in short read assemblies. On top of that, the combination of long read assemblies and custom TE libraries maximise the representation and annotation of the transposable element content as highlighted here for the genome of *H. glaber* (Figure 2). The first TE annotation of *H. glaber* without a

custom TE library showed a repetitive content of about 25% [47] but the use of a proper TE library increased the content to ~34% (**Figure 2**) and, finally, the employment of long reads increased the total content to 38% (**Figure 2**, **Table 1)**.

We then proceeded with a more detailed analysis of the TE accumulation over time for rodents and bats to investigate possible differences in accumulation between short- and long-lived species and to find shared features between long-lived species. The main difference between short- and long-lived species of rodents is represented by a stark drop in non-LTR retrotransposon accumulation at recent times (0-5% divergence; **Figure 3**). Similarly, the long-lived species of bats show a drop in non-LTR retrotransposon accumulation at recent times (**Figure 4**) while presenting an overall accumulation of class II transposons (DNA transposons and Helitrons). Previous studies hypothesised that bats have a higher tolerance for the activity of transposable elements with alternative ways to dampen potential health issues due to this activity [41, 56]. Given our observation about the common drop of non-LTR retrotransposon accumulation in bats and naked mole rat, we add to the aforementioned hypothesis that it is the particular repression of the activity of non-LTR retrotransposon that enhances the resistance to cancer. The non-LTR retrotransposons are the most present types of TEs in rodents (**Figure 3**, **Table 1**) 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 the long-lived and short-lived species considered in this study (**Figure 5**, **Figure 6**) using the density of recent insertions (DI). The naked mole rat showed the lowest DI of non-LTR retrotransposons followed by *C. porcellus* that has a lower cancer incidence than *M. musculus* and *R. norvegicus* [55] (**Figure 5**). Given the pattern of DI observed in rodents, the recent accumulation of non-LTRs appears to be related to the lifespan

of these species (**Figure 1a**, **Figure 5**, **Figure 7a**). On top of that, the genome of the naked mole rat does not contain intact ORFs of LINE retrotransposons which is another clue for the absence of currently active non-LTR retrotransposons (**Figure 5**). The absence of intact LINE ORFs in the long read genome assembly confirms the same observation previously reported on the short read based genome assembly [60]. In bats, we observed that *M. molossus*, the bat species with the shortest lifespan, showed the highest DI, though no intact LINE ORFs were found in its genome. The high DI in *M. molossus* genome (**Figure 6**) is driven by a striking accumulation of non-autonomous SINEs. SINEs are about 300 bp long and do not present any protein coding sequence which make them unable to retrotranspose autonomously. Indeed, SINEs rely on the exploitation of the retrotransposition machinery provided by LINEs to move throughout the genome (**Figure 7**). Since *M. molossus* genome does not present any intact LINE ORFs, it is reasonable to consider both LINEs and SINEs to be inactive in this genome (**Figure 7b**).

Theoretically, the absence of recent and active non-LTR retrotransposon insertions could make the genome more stable and reduce the frequency of cancer incidence. Nonetheless, the recently inserted SINEs and LINEs 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 that can be 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 [61]. Among the killifishes, the turquoise killifish has a lifespan of 4-8 months

(likely the shortest lifespan among vertebrates [61]) but despite the accumulation of mutations that leads to a fast ageing, the species is fully adapted to his ecological niche demonstrating that a rapid ageing is not necessarily linked to reduced fitness (**Figure 7b**). In addition to being considered genomic mutations, SINEs can also exert physiological effects by being transcribed. For example, in humans, SINEs of the Alu family have been found to be involved in the inflammatory pathway: upon viral infection, the interferon 1 mediates the overexpression of Alus which function as an amplifier of the immune response by stimulating the intracellular viral RNA sensor systems [62]. Despite, the 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 [63–65]. Furthermore, as non-coding RNAs are emerging as essential components of neuroinflammation, a study has included Alu SINEs as possible promoters of neurodegenerative diseases [66]. More in general, the increased transcription of non-LTR retrotransposons (both LINEs and SINEs) in humans may contribute to the so-called "sterile inflammation" (**Figure 7b**), 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") [67]. Therefore, it is possible that the great accumulation of SINEs in *M. molossus* trigger phenomena similar to the sterile inflammation and inflammageing that cause a shortening of its lifespan. For these reasons *M. molossus* may be an exceptional model for biogerontology.

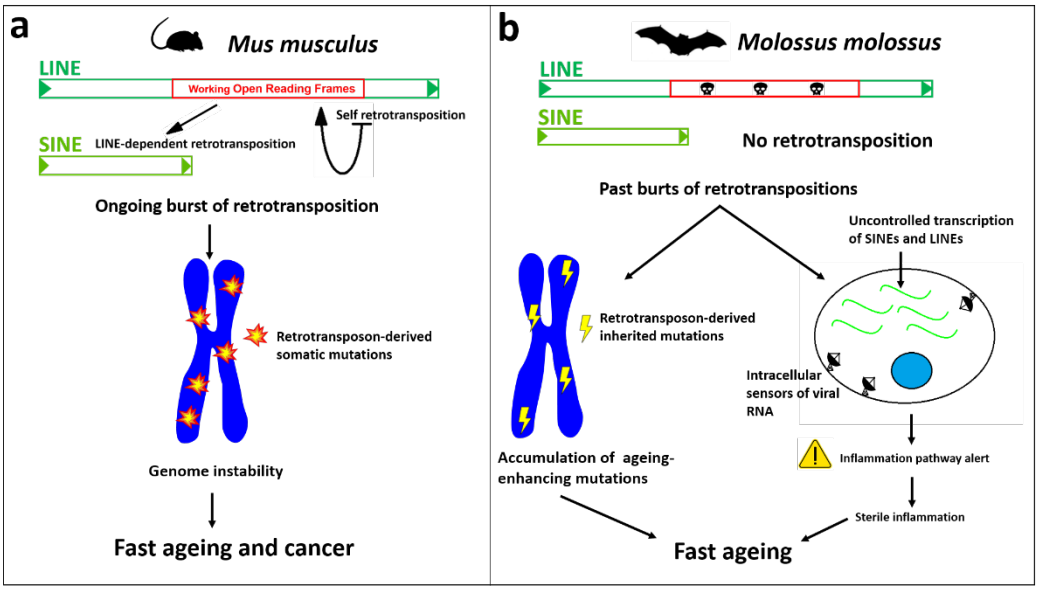

Figure 7. Hypothetical contribution of non-LTR retrotransposons to the fast ageing of rodents and bats. **a**) *Mus musculus* genome bears active non-LTR retrotransposons that destabilize the genome with continuous events of retrotransposition causing cancer and other diseases that reduce the mouse lifespan. **b**) *Molossus molossus* is both a cancer-resistant and short-lived species. The high density of non-LTR retrotransposons derives from a past accumulation of insertions (**Figure 4**) that may have resulted in the accumulation of ageing enhancing mutations and the continuous transcription of these retroelements may trigger cellular inflammation and cause a widespread sterile inflammation. In this scenario, the fitness of *M. molossus* is not impaired but its lifespan is reduced.

Conclusions

An increasing number of studies are finding a link between TEs, ageing, and cancer. As discussed in previous studies [39, 41], it is reasonable that genome instability derived by TEs could facilitate age-associated pathologies. In this work, we showed a relationship with retrotransposons activity and cancer incidence in rodents. The only cancer-resistant rodent species (naked mole rat) has a genome devoid of active LINEs and the lowest density of recently inserted non-LTR retroelements (SINEs and LINEs). The six analyzed species of bats are all cancer resistant, but they show widely different lifespans (from 37 years *Myotis myotis* to less than 6 years *Molossus molossus*). This particular taxon of mammals shows, together with a general cancer resistance, a drop of accumulation (and/or activity) of non-LTR retrotransposons in recent times similarly to the naked mole rat. The similar drop in non-LTR

retrotransposon accumulation may indicate a possible involvement of retrotransposons in the incidence of cancer. The fast-ageing *M. molossus* has the highest density (DI) of non-LTR retroelements among bats (**Figure 6**) driven by the accumulation of SINEs. Despite the higher accumulation of SINEs in *M. molossus*, the absence of intact LINEs (that mobilise SINEs) makes SINEs inactive and not a source of new somatic mutations through their retrotransposition. This genomic feature makes *M. molossus* theoretically as cancer resistant as the other bat species. In humans and other mammals non-LTR retrotransposons are also part of the inflammation pathway [REF]. The high density of SINEs observed in *M. molossus* may promote ageing trough their high mutation load and their pervasive transcription may participate to the processes similar to the sterile inflammation and inflammageing. In conclusion, the non-LTR retrotransposon content may act as promoting factor, more than as simple biomarker, in cancer evolution and other age-associated conditions.

Methods

Samples

In order to avoid biases related to potential underestimations of the TE content, in this study we chose species of mammals from which high-quality genome assemblies were available. The 10 species taken in to account and the accession numbers of their assemblies are listed here: *Mus musculus* GCA_000001635.8 **[**Genome Reference Consortium mouse reference 38], *Rattus norvegicus* GCA_000001895.4 and [68], *Cavia porcellus* GCA_000151735.1[69], *Heterocephalus glaber* GCA_014060925.1 [48], *Rousettus aegyptiacus* GCA_014176215.1 [70] , *Rhinolophus ferrumequinum* GCA_014108255.1 [70], *Molossus molossus* GCA_014108415.1 [70] *Pteropus vampyrus* GCA_000151845.1 [69], *Myotis myotis* GCA_014108235.1 [70], *Myotis lucifugus* GCA_000147115.1 [69]. The species selected are from Rodentia and Chiroptera orders. These two orders have been separated in the Mesozoic

era 90 mya ago [71]. The large evolutionary distance among these two orders allowed us to find genomic features potentially conserved in all mammals.

Body mass is generally positively related to the lifespan [45] therefore, in our study, we chose to analyse genomes belonging to small-sized mammals (between 12 g and 1.1 kg). The comparisons were carried out between species belonging to the same taxonomical order to maximize their shared evolutionary history, physiological features and life history traits, then the species were also compared on the basis of their lifespans (**Figure 1**). The genome assemblies analysed are all based on Sanger or PacBio technologies to maximize their assembled repetitive content [72–75]. The short (GCA_000230445.1 [47]) and long read based genome assemblies of *H. glaber* were compared in order to estimate the differences in TE content due to technological biases.

Repetitive element libraries and annotation

To analyze the content of transposable elements in a genome, the genome assemblies must be annotated with proper libraries of TE consensus sequences. A consensus represents the reference sequence of a specific subfamily of transposable elements. The repeat annotation softwares (like RepeatMasker) make use of such library of consensus sequences to find instances (or insertions) of each repeat in the genome assemblies. From the alignments between the insertions found in the genome assembly and their consensus sequences, it is possible to estimate the age of such insertions on the basis of the genetic distance given by the alignments: the higher the distance, the higher the age of the insertions.

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 [76] The resulting consensus sequences were merged with the ancient mammalian TE families (LINE2 and MIR SINEs) downloaded from RepBase (https://www.girinst.org/repbase). All the genome assemblies of rodents (but *H. glbaer*) were masked using RepeatMasker (v. 4.1.0) and the Rodentia-specific library (Repbase release 20181026). Similarly, the bat genome assemblies were masked using the Chiroptera-specific library from Repbase merged with the curated TE libraries produced by Jebb et al. [70]. All RepeatMasker annotations were performed using the options -a -xsmall -gccalc -excln. The TE annotations were then visualized through barplots (landscapes) that show in the X-axis the genetic distances from the consensus sequences (calculated as Kimura 2-p distance [49]) and in Y-axis the percentage of the genome occupied by each TE category. The barplots were made using the ggplot2 R package [77]. For each species, the genome assembly size and the percentage of each TE categories listed in **Table 1**, **Table 2** and **Figure 2** were retrieved from the table files produced by RepeatMasker.

Density of young non-LTR retrotransposons

To evaluate the density of recent non-LTR retrotransposon insertions (LINEs and SINEs), we selected all the non-LTR retrotransposon hits found by RepeatMasker with a divergence from consensus lower than 3%. Then we calculated the density of insertion (DI) [12] for each species as the ratio between the number of recent insertions and the corresponding assembly size expressed in Gigabases:

$$Density\ of\ insertion\ (DI) = \frac{Number\ of\ insertions}{Genome\ size\ (Gb)}$$

Intact ORF detection

To check the presence of intact LINE retrotransposons (therefore potentially active) we looked for complete open reading frames (ORFs) that encode for the enzymatic machinery used by

LINEs to retrotranspose themselves and the non-autonomous SINEs. In each species the sequences of LINEs were with BEDTools getfasta using the LINE coordinates annotated by RepeatMasker. A fasta file for each species was produced and used as input for the R script orfCheker.R (https://github.com/jamesdgalbraith/OrthologueRegions/blob/master/orfChecker.R) to find intact LINE ORFs. The script considers a LINE ORF to be intact if it contains both complete reverse transcriptase and endonuclease domains.

Acknowledgments

This study has been funded by MAPS Department funds BIRD2021 - prot. BIRD213010, University of Padova.

References

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. Peto R. Epidemiology, multistage models, and short-term mutagenicity tests. Int J Epidemiol. 2016;45:621–37.

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

44. 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.

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

46. 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. Mol Ecol Resour. 2021;21:263–86.

47. 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.

48. 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.

49. Kimura M. Journal of Molecular Evolution: Introduction. J Mol Evol. 1980;16:111–20.

50. 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.

51. 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.

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

53. 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.

54. Eunjung L, Rebecca I, Lixing Y, Omer G, Psalm H, J. LL, et al. Landscape of Somatic Retrotransposition in Human Cancers. Science (80-). 2012;337:967–71.

doi:10.1126/science.1222077.

55. 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.

56. 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.

57. Foley NM, Hughes GM, Huang Z, Clarke M, Jebb D, Whelan C V., et al. Growing old, yet staying young: The role of telomeres in bats' exceptional longevity. Sci Adv. 2018;4.

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

59. 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.

60. 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.

61. 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.

62. 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.

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

64. Lagisquet J, Zuber K, Gramberg T. Recognize Yourself—Innate Sensing of Non-LTR Retrotransposons. 2021;:1–19.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, et al. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science (80-). 2011;334:521–4.

72. 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.

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

74. Armitage D. Rattus norvegicus. Animal Diversity Web. Accessed October 27, 2020.

2004;:https://animaldiversity.org/accounts/Rattus_norveg.

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

76. 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.

77. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. https://ggplot2.tidyverse.org.