Title

Implications of the reserve-capacity hypothesis for in-bred laboratory rats: a cross-species comparison of telomere content in consideration of experimental confounds

Abstract

The genetic, anatomical, and physiological similarities of mice and rats to humans have made rodent models enormously valuable in biomedical research. The large demand for lab mice (*Mus musculus*) has motivated breeding centers to maximize reproductive output by promoting the most philoprogenitive mouse strains, which artificial selection pressure appears to have elongated their telomeres. The reserve-capacity hypothesis explains this shift by positing that a species' telomere length is optimized in response to selection against senescence, resulting in longer tissue-repairing telomeres when strong and in shorter tumor-suppressing telomeres when weak. The elongation of captive-mouse telomeres purchases robust reproductive viability at the cost of increased tumor susceptibility, which when safety testing with such models likely results in overestimation of carcinogenicity and underestimation of tissue damage. Laboratory rats are preferred over mice when experimental design requires a more expansive behavioral repertoire (i.e. higher intelligence), as is the case, for example, in self-administration models of substance abuse. The Brown Norway rat (*Rattus norvegicus*) is an in-bred strain like *Mus musculus*, so there has been similar opportunity for artificial selection pressure to have hyperextended their telomeres. This study compared the telomere content of three species of rats, four species of mice, a few other rodent species (vole, capybara, rabbit), as well as human, and determined that the lab rat’s telomeres are substantially longer, by length and proportion of genome, than those of all the other species listed except for the African grass rat (*Arvicanthis niloticus*). The results indicate that while *R. norvegicus* telomeres are not longer than those of all their wild counterparts, they may be sufficiently hyperextended to invalidate them as models in studies measuring carcinogenicity and tissue damage. Future studies could investigate the causal relationship between artificial selection in breeding programs and *R. norvegicus* telomere length.

Keywords

telomeres, laboratory rat, reserve-capacity hypothesis

Background

Telomeres are tandem (TTAGGG)n repeats at the ends of linear chromosomes which stabilize the genome and enable replication. Evidence from hybridization of biotinylated oligonucleotides of this sequence suggests conservation across mammals, birds, amphibians, reptiles, and teleosts, of not only the telomeric sequence but also the chromosomal location (1). Mammalian telomeres are associated with a six-member shelterin protein complex which protects the ends of chromosomes from DNA damage machinery by the formation of a lariat-like t-loop. Within shelterin the Pot1-TPP1 heterodimer, the telomere-binding proteins TRF1

and TRF2, and the interacting factors Rap1 and Tin2 work in concert to protect the chromosome (2). Recognition by the DNA damage machinery as broken DNA would result in end-joining, repair or recombination that may destabilize chromosomes (3). Progressive attrition of telomeres occurs in all proliferative cells due to the incompleteness of lagging strand synthesis, exonucleolytic processing events, oxidative damage, and other factors. When telomeres reach a critical length, the cell cycle is arrested and the cell enters replicative senescence, in which state it no longer divides, which links telomeres to longevity and regenerative capacity (4).

The regenerative capacity of stem cells is contingent on maintaining telomere length (5). Telomeres are also responsible for ensuring correct chromosome segregation during mitosis, and silencing certain genes close to the telomeres (a phenomenon termed subtelomeric silencing) (2). To compensate for the gradual telomere attrition induced by the 'end replication problem,' i.e. the inability for DNA polymerases to synthesize the ends of linear chromosomes, the enzyme telomerase affixes TTAGGG repeats to the ends of chromosomes. Pluripotent stem cells exhibit robust telomerase expression. Developing embryos do as well, but telomerase activity in adult cells is not sufficient to countervail the telomere degradation associated with ageing. Short, dysfunctional telomeres have been linked to rare human genetic diseases such as idiopathic pulmonary fibrosis, dyskeratosis congenita, and aplastic anemia. Another common feature of these diseases is the impairment of tissue regeneration, a function in which telomeres have been implicated (2,6). Studies of telomerase-deficient mice (TERC-knockout mice; TERC is a component of telomerase) have shown that limiting telomerase activity likewise limits longevity: progressive generations of these mice showed shorter telomeres and lifespans (2).

Telomere shortening is programmed in human development, occurring in most cells – with the exception of certain stem cells. This is due to the downregulation of telomerase activity via TERT silencing. Telomerase reverse transcriptase (TERT) affixes GGTTAG repeats to the 3' end of chromosomal DNA to counteract telomere attrition. The suppression of telomerase activity in somatic cells and the barrier to proliferation that this imposes indicates a tumor-suppressing pathway to limit tumorigenic overgrowth (following a delay) (7). This anti-cancer adaptation, however, comes to have the opposite effect in nascent cancer cells without the cell cycle arrest pathways, as telomere dysfunction destabilizes the genome towards a telomere crisis, escape from which requires telomerase activity. A cell transformed in this way will have a stable but considerably rearranged genome with potentially tumorigenic mutations (7).

Telomeres degrade with each cell division in the absence of telomerase, so longer telomeres would benefit the organism by raising the number of remaining divisions available to a cell or cell line before the induction of senescence (this is referred to as *reserve capacity*); there is, however, a trade-off, namely that with augmented proliferative capacity comes heightened tumor susceptibility, as the same mechanism that enables tissue repair by cell division would also enable neoplastic

proliferation. The reserve-capacity hypothesis posits an antagonistic pleiotropy to explain the proliferative limits of somatic cells: a species' telomere length is optimized based on the selection against senescence that it experiences. When externally induced mortality is high, the priority remains rapid reproduction, and tumor development is a problem that few members of the species live long enough to suffer. When selection against senescence is weak, then it behooves the population to set a tumor-suppressing fail-safe in the form of shorter telomeres, as extending an iteroparous lifespan provides more opportunity for reproduction. This antagonistic pleiotropy results in long telomeres enhancing resistance to tissue damage but increasing risk of tumor growth, and in short telomeres producing the opposite effect, with each species landing at an optimum based on its life cycle and reproductive strategy (6).

Lab mice are widely used in biomedical research due to the important genetic, anatomical, and physiological similarities they share with humans (8). Genomic telomere content, however, may represent an important difference: human telomeres average about 10 kilobases in length in blood cells (9), while those of laboratory mice can range from 30 up to 150 kb (9,10). It has also been observed that laboratory mice have substantially longer telomeres than their wild counterparts (10), which suggests an effect of artificial selection pressure in breeding programs: selection for productive mice (and no selection against senescence) offsets the optimal balance between tumor suppression and tissue repair, resulting in robustly philoprogenitive mouse strains that have heightened neoplastic vulnerability (6).

Although the mouse may be more popular, the laboratory rat *R. norvegicus* was the first mammalian species domesticated for scientific investigation, and currently there are many inbred strains in use for research (11). Humans and rodents diverged about 80 million years ago, and the last common ancestor of the rat and mouse is estimated to have lived 12-24 million years ago. The rat genome (2.75 gigabases, Gb) is smaller than the human genome (2.9Gb) but larger than that of the mouse (~2.6Gb). All three genomes encode similar numbers of genes, the majority of which have persisted without duplication or deletion since the last common ancestor. Intronic regions are also highly conserved. Almost all genes associated with human disease find orthologues in rats, although their rates of synonymous substitutions vary from the others. About 30% of the rat genome aligns only with that of the mouse, of which a considerable proportion is rodent-specific repeats (11).

While the issue of potentially confounding hyper-long telomeres (as is found in mice) is unlikely to be as great a problem in out-bred strains of lab rat like the Sprague-Dawley (12), the inbred Brown Norway rat might have been subjected to similar selective pressure as *M. musculus* and thus the length of its telomeres may raise concerns in certain studies. Adaptations to increase early life fitness may result in accelerated age-related somatic decline, and compromise *R. norvegicus* as a model for human susceptibility to cancer and tissue damage.

Results

When estimating telomere length from whole genome sequence data, a threshold (k) for the minimum number of TTAGGG repeats must be established to classify the segment as telomeric. Ding et al. and Nersisyan and Arakelyan use a k of 7 (13,14), and in a measurement of absolute telomere length from qPCR, Montpetit et al. use a standard that is 14 repeats long (15). In this study, telomere lengths were calculated and comparisons were made based on two thresholds of k = 7 and k = 14.

Telomere content as proportion of genome (total telomeric segments in bp divided by total length of genome in bp) and as absolute length (in kb) was determined for 11 species at two k values. Figure 1 shows the telomeric percentage of genome for all species. Because the telomere content of *R. norvegicus* and especially that of *A. niloticus* tower over the others, a separate inner panel portrays the telomeric proportion of genome for the 9 species without the outliers. As depicted in Table 1, the range at k = 7 extends from *P. leucopus*' 0.0003 % to *A. niloticus*' 0.11 %, covering three orders of magnitude. Measuring telomeres by length, as opposed to proportion, does not affect the relationships depicted in Figure 1, with Figure 2 being almost indistinguishable but for the labels. The range of total telomeric length of genome extends from *P. leucopus*' 7.428 kb to *A. niloticus*' 2303 kb, again across multiple orders of magnitude.

Figures 1 and 2 reveal that for species with high telomere content like *A. niloticus*, elevating the threshold k from 7 to 14 does not greatly impact the measurement: the African grass rat's telomeric proportion of genome drops from 0.114% to 0.107% (still about 93% of its low-threshold value) and its total telomeric length declines from 2302 to 2172 kb (again, a reduction of about 7%). Looking at the same measurements for the rabbit *O. cuniculus*, however, reveals reverse patterns: the telomeric proportion of the rabbit genome falls from 0.00094% to 0.000067%; and its total telomeric length, from 21 kb to 1.5 kb, as k rises from 7 to 14 – both of these shifts representing a reduction of approximately 93%. It seems that for organisms without exceptionally long telomeres, adjusting the minimum repeat threshold can affect telomere measurement much more drastically than it does in long-telomere genomes. In the case of *P. leucopus* and *H. hydrochaeris*, this increase in k value results in a report of zero telomere content, and so these species are excluded from Tables 3 and 4 which present values measured at k = 14. The forced exclusion of these species at this threshold suggests that a threshold value of 14, while interesting to consider in an analysis like this, may be too high to represent a true telomeric segment of DNA, as it would lead us to conclude that the white-footed mouse and capybara have no telomeres at all, which is clearly not the case.

The telomere content by chromosome for the Brown Norway rat, the black rat, and the African grass rat are depicted in Figures 3 and 4. These three rat species appear to occupy three different strata of telomere content, with the African grass

rat having the highest values; the black rat, the lowest; and the laboratory rat, landing somewhere in the middle. An exception lies on chromosome 14, the telomeric proportion of which is much higher for the lab rat than for the grass rat. When comparing absolute length in Figure 4, however, this exception fails to appear. The k criterion does not appear to affect the relationships between telomeric content measurements, their relative values remaining similar after the increase from 7 to 14.

The relationship of the laboratory mouse telomere content compared to wild counterparts was also examined at the chromosome level. Figure 5 depicts the telomeric proportion by chromosome for three species of mouse: *M. spretus*, *M. musculus*, and *Peromyscus maniculatus bairdii*. Unlike the rats, the relationships between the relative telomere content of *M. spretus* and *M*. musculus vary appreciably between methods of measurement, as well as between threshold values. The telomeric proportion of the laboratory mouse genome at k = 7 was measured to be 0.0050 %, while at k = 14 this number drops to 0.0034 %. This is a slightly larger dip than the telomeric proportion of the laboratory rat genome takes, from 0.040 % to 0.030 % as k increases from 7 to 14. *Mus spretus* 'loses' about half of its telomere content when the threshold is raised, with a telomeric percentage of 0.012 % at k = 7 dropping to 0.0065 % at k = 14. The differences between *M. spretus* and *M. musculus* in Figure 5, for the first 10 chromosomes at least, drop significantly when the threshold is raised, suggesting that while the white-footed mouse has higher telomere content overall, the genome of the laboratory mouse has a larger proportion of its telomeres contained in segments of at least 14 repeating units of TTAGGG.

Discussion

An ideal model organism for biomedical study should obviously share basic attributes with the species which we cannot subject to such study (humans) but which we care to know about. Given the relationship between telomere length, tumor vulnerability and somatic tissue repair capacity, the comparatively high telomere content of in-bred laboratory rodent models should be carefully considered in studies measuring carcinogenicity and tissue damage. Table 1 shows that measuring telomeric proportion of genome at a threshold (k) of 7 tandem TTAGGG repeats, the laboratory mouse and rat contain 4.7 and 37.7 times the telomere content of *H. sapiens*, respectively. Measuring in kilobases, these numbers change little (4.1 and 32.1). These ratios reveal significant differences in telomere content between the species of interest, *H. sapiens*, and model organisms, *M. musculus* and *R. norvegicus*. Weinstein and Ciszek considered the telomeric hyperextension of the lab mouse telomeres a confounding problem (6), and given that the genome of *R. norvegicus* contains 3.3 times more telomeres (by kb, k = 7) than *M. musculus*, it is reasonable to suspect that the confound may be even greater in studies of the laboratory rat that involve telomere-relevant phenomena.

Peculiarly, both the laboratory rat and mouse lie intermediately in telomere content between their wild relatives that were investigated in this study. As seen in Figures 3 and 4, the telomere content by chromosome of *R. norvegicus* is generally much higher than that of *R. rattus*, but much lower than *A. niloticus*. Per Table 2, *R. norvegicus* has about ten times the total telomere length of *R. rattus*, but only 46% as much as *A. niloticus*. It seems appropriate that the African grass rat is not commonly employed in laboratory studies, as the proportion of its genome at k = 14 is a staggering 262 times that of *H. sapiens*. Based on the reserve capacity hypothesis, this kind of telomere content would suggest very weak selection against senescence in the *A. niloticus* population.

As seen in Figures 5 and 6, *M. musculus* tends to find the telomere content of a given chromosome somewhere between *P. bairdii* and *M. spretus*, though generally closer to the latter. Table 1 shows that the telomeric proportion of the laboratory mouse genome is about 12 times that of *P. bairdii* but only 41% that of *M. spretus*. Like *R. norvegicus*, *M. musculus* is not the species with the most potentially confounding telomere content, but it remains significantly higher than that of *H. sapiens*, and the other species of rodent (European water vole, rabbit, and capybara).

The laboratory rat *R. norvegicus* contains more telomere content than all other species except for the African grass rat *A. niloticus*. Next in line is the Algerian mouse (*M. spretus*), and then the laboratory mouse claims fourth place in this running. Compared to the other species of rodent in this study, the telomere content of laboratory animals is very high (see Tables 1-4 for ratios of telomere content between each pair of species). Each of these laboratory animals, however, is outdone or even eclipsed by one of their wild counterparts, which illustrates the high inter-species variability of telomeres. Some strains of rat, like the Sprague-Dawley, exhibit such high variability that even within the same breeding program, rodents from different litters can be significantly genetically distinct. While sometimes it is advantageous to rely on the genetic consistency of in-bred strains, the variability of out-bred rodent models might help to avoid such confounds as hyperextended telomeres (12).

This study was limited by the number of publicly available rodent genomes sequenced at the chromosome level. When more species of rat, mouse, and rodents in general are represented this way in NCBI's databases, more comprehensive comparisons can be made to assess the potential for experimental confounds, as well as the effect of breeding programs on telomere length. There is also the inevitable arbitrariness of the k value, which was determined in this study from precedents in the literature (13,14,15), but which could benefit from more solid theoretical justification. Whether there is something biologically meaningful, or more inherently telomeric, about a stretch of at least 7 or 14 (as opposed to 3 or 10, etc.) repeats of TTAGGG remains to be seen. Moreover, as previously stated, two species in this study (the white-footed mouse and capybara) had to be excluded from analyses at k = 14 because their genomes contain no stretches of at

least 14 telomeric repeats. Finally, while the reserve-capacity hypothesis presents a compelling explanation for the length of the lab mouse telomeres compared to their wild counterparts, no conclusions of such causation can be drawn from this study. The lab rat telomere data is consistent with this hypothesis, and it is possible that artificial selection in breeding programs has led to the elongation of both *M. musculus* and *R. norvegicus* telomeres, but neither of these species represent the most telomere-heavy mouse or rat in general, and the fact that their telomere content is higher than most of the species surveyed here does not suggest anything about a mechanism for this difference.

Conclusion

Lab rat telomeres are longer than not only those of humans but also the lab mouse, whose telomere length has already been flagged as potentially problematic. As telomeres appear to affect the number of remaining divisions available to a cell or cell line (reserve-capacity), studies that measure somatic insult and tumorigenesis with rodent models would likely limit telomere-based confounds by deploying an out-bred rodent species, or at least one with telomere content more comparable to the biomedical species of interest, *H. sapiens*. If the reserve-capacity hypothesis is correct, then the severity of tissue damage in these studies is likely to be underestimated, while the extent of carcinogenicity would probably be magnified, as the model organisms *M. musculus* and *R. norvegicus* have a loftier Hayflick limit than human cells, with more propensity for both productive and unregulated division (6).

Future studies could investigate more species of rat and mouse for a more comprehensive understanding of relative telomere content. In particular, the telomeric differences between in-bred Brown Norway rat strains and the wild Asian house rat from which this strain descends, could begin to expose the causal relationship between breeding programs and telomere hyperextension. Finally, many studies of telomere length would benefit from a theory-based optimal k value for standardizing their measurement, i.e. what ‘counts’ as telomeric in a sequence.

Methods

The aim of this study was to compare the telomere content of laboratory rodent models with their wild counterparts and with humans, to investigate the concerns of Weinstein and Ciszek (6) regarding the confounding effect of hyperextended telomeres on studies measuring tissue damage and carcinogenicity. A python script (genomics_final_project.py) was used to access and process the data, as well as prepare figures and tables (except for Table 5). The species investigated here were the lab rat *Rattus norvegicus*, the black rat *Rattus rattus*, the African grass rat *Arvicanthis niloticus*, the lab mouse *Mus musculus*, the Algerian (or Western Mediterranean) mouse *Mus spretus*, the white-footed mouse *Peromyscus leucopus*, the prairie deer mouse *Peromyscus leucopus bairdii*, the European water vole *Arvicola amphibius*, the rabbit *Oryctolagus cuniculus*, the capybara *Hydrochoeris hydrochaeris*, and the human *Homo sapiens*.

The Entrez system is the National Center for Biotechnology Information's (NCBI) primary interface for searching and retrieving text from not only the PubMed database of biomedical literature, but also 38 other databases containing information about genome, gene, protein sequence, structure, genetic variation, and gene expression. The database of interest to this study was the Assembly resource for access to genome assemblies, including both submitted data and NCBI RefSeq assemblies. With versioned accession numbers for sequences, accessing the Assembly database through Entrez provides direct downloads of nucleotide sequences for genome analysis (16).

Species were selected based on data availability and relevance to the aim of the study. Two wild counterparts for the lab rat (*Rattus rattus* and *Arvicanthis niloticus*) and for the lab mouse (*Mus spretus* and *Peromyscus leucopus bairdii*) provided per-chromosome comparisons for Figures 3-6. The telomere contents of these species were intended to represent the 'wild' versions of the laboratory rodents to elucidate the potential effect of artificial selection in breeding programs. Though they did not suffice as perfect controls, they did lead to some interesting comparisons that revealed at least one species of mouse (*Mus spretus*) and rat (*Arvicanthis niloticus*) have extremely high telomere content for which no breeding program is to blame. The water vole, rabbit, and capybara were included to compare laboratory animal telomere content to other rodents.

NCBI contains chromosome-level whole-genome sequence data for all eleven species investigated here (see Table 5 for Assemblies and accession IDs). Each species was represented in the python script as a Species object, which contained Seq objects that held the relevant telomeric data. Nucleotide sequences in fasta format were accessed from the Species objects using Entrez and temporarily loaded into Seq objects, which used the finditer function from the regular expression (re) module to locate all instances of at least one telomeric unit: TTAGGG. The indices of all these instances were stored as a list of tuples and written to files. This way, the data could easily be reprocessed with a different k value, which would have required more time if Entrez had to be deployed again.

To calculate the total telomeric proportion of genome, first the length of each sequence was accessed from the string of the SeqIO sequence record object fetched by Entrez. Then, the difference of the stop and start locations in each tuple of indices of $(TTAGGG)_n$ was divided by 6 (the length of the telomeric unit in base pairs) and evaluated with respect to k. If greater than or equal to the threshold, the segment classified as telomeric and was summed with all the others that met the criterion. The sum of all telomeric segments in base pairs divided by the length of the sequence determined the telomeric proportion of that sequence (i.e. chromosome), and summing all the telomeric segments from each chromosome and dividing by the sum of the lengths of each chromosome resulted in the telomeric proportion of the genome for that species.

References

1. Meyne J, Ratliff R, Moyzis R. Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Evolution. 1989;86:7049-53.
2. Donate L, Blasco M. Telomeres in cancer and ageing. Philosophical Transactions of the Royal Society. 2011;366:76-84.
3. Shekhidem H, Sharvit L, Leman E, Manov I, Roichman A, Holtze S, Huffman D, Cohen H, Hildebrandt T, Shams I, Atzmon G. International Journal of Molecular Sciences. 2019;20(13):3233
4. Shay J, Wright W. Telomeres and telomerase: three decades of progress. Genetics. 2019;20:299-309
5. Morgan C, McCartney A, Donoghue M, Loughran N, Spillane C, Emma C, Teeling E, O'Connell M. Molecular adaptation of telomere associated genes in mammals. BMC Evolutionary Biology. 2013;13:251
6. Weinstein B, Ciszek D. The reserve-capacity hypothesis: evolutionary origins and modern implications of the trade-off between tumor-suppression and tissue-repair. Experimental Gerontology. 2002;37:615-27.
7. Maciejowski J, Lange T. Telomeres in cancer: tumour suppression and genome instability. Molecular Cell Biology. 2017;18:175-87.
8. Bryda E. The Mighty Mouse: the impact of rodents on advances in biomedical research. Missouri Medicine. 110(3):207-11.
9. Kipling D, Cooke H. Hypervariable ultra-long telomeres in mice. Nature. 1990;347:400-2.
10. Hemann M, Greider C. Wild-derived inbred mouse strains have short telomeres. Nucleic Acids Research. 2000;28:22
11. Pachter L. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature. 2004;428:493-521.
12. Gileta A, Fitzpatrick C, Chitre A, St. Pierre C, Joyce E, Maguire R, McLeod A, Gonzales N, Williams A, Morrow J, Robinson T, Flagel S, Palmer A. Genetic characterization of outbred Sprague Dawley rats and utility for genome-wide association studies. doi: https://doi.org/10.1101/412924
13. Ding Z, Mangino M, Aviv A, Spector T, Durbin R. Estimating telomere length from whole genome sequence data. Nucleic Acids Research. 2014;42(9):e75. doi: 10.1093/nar/gku181
14. Nersisyan L, Arakelyan A. Computel: Computation of Mean Telomere Length from Whole-Genome Next-Generation Sequencing Data. PLoS One. 2015;10(4):e0125201. doi: 10.1371/journal.pone.0125201
15. Montpetit A, Alhareeri A, Montpetit M, Starkweather A, Elmore L, Filler K, Mohanraj L, Burton C, Menzies V, Lyon D, Collins J, Teefey J, Jackson-Cook C. Telomere Length: A Reivew of Methods for Measurement. Nurs Res. 2014;63(4):289-299. doi:10.1097/NNR.0000000000000037
16. NCBI. Entrez Help. 2005. https://www.ncbi.nlm.nih.gov/books/NBK3837/#_NBK3837_pubdet_

Figures

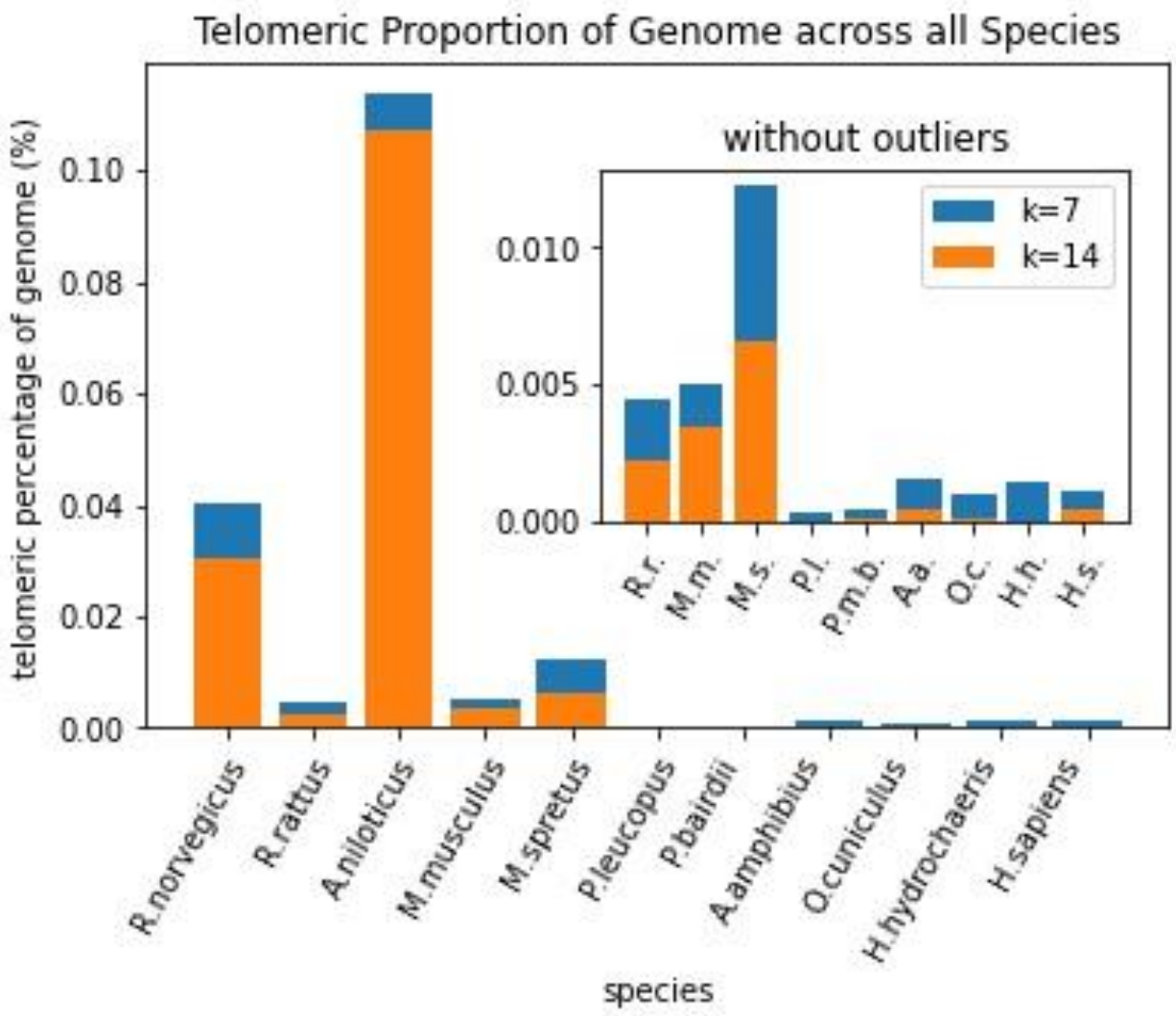

Figure 1:

Telomeric percentage of genome calculated from total base pairs contained in segments of at least k tandem repeats of TTAGGG divided by genome length x 100%. The telomere content of *R. norvegicus* and *A. niloticus* dwarf those of the other species, which are presented again separately on a smaller scale in the inner panel.

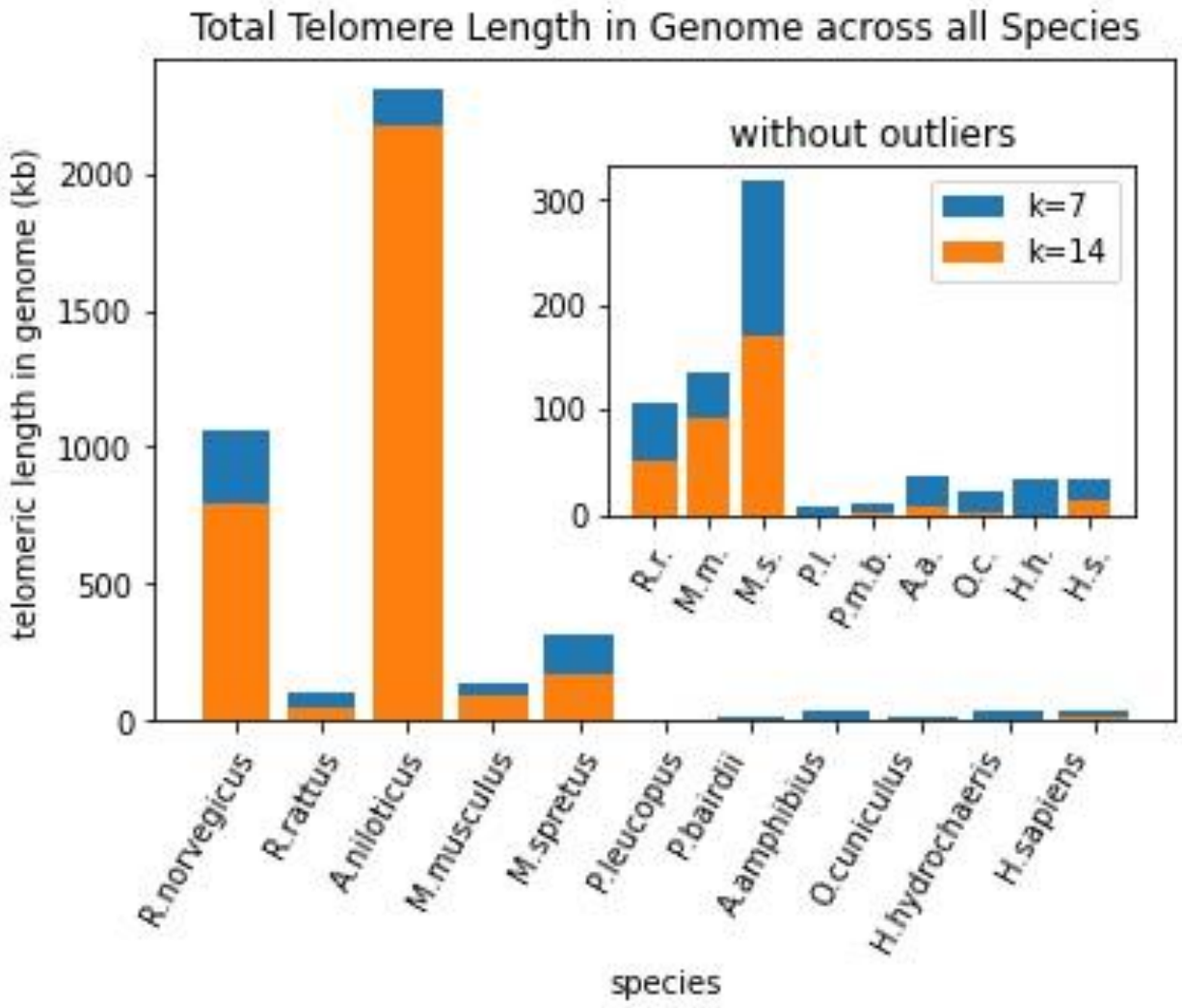

Figure 2:

Total telomeric length of genome calculated from total base pairs contained in segments of at least k tandem repeats of TTAGGG. The telomere content of *R. norvegicus* and *A. niloticus* dwarf those of the other species, which are presented again separately on a smaller scale in the inner panel.

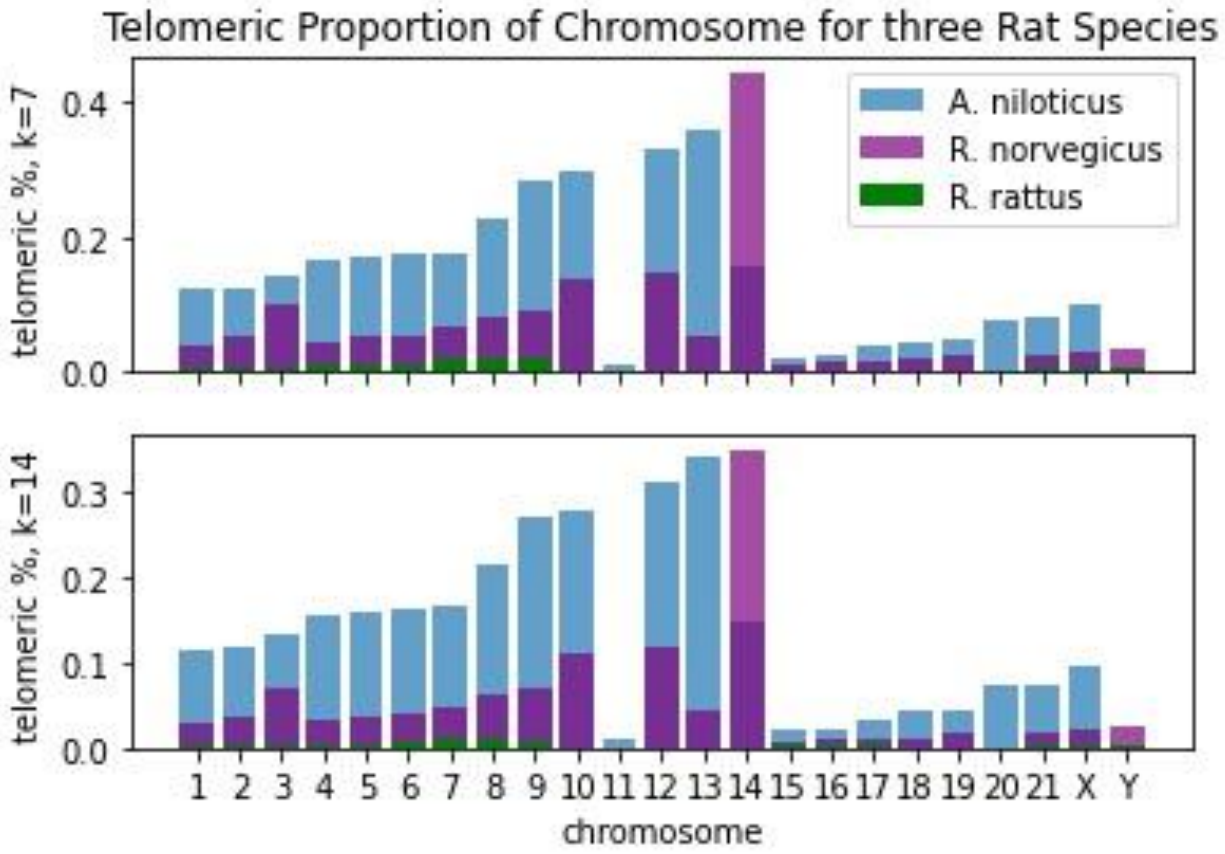

Figure 3:

Telomeric percentage of each chromosome for three rat species, calculated from total base pairs contained in segments of at least k tandem repeats of TTAGGG divided by length of the chromosome x 100%. Not all species have the same number of chromosomes: *R. norvegicus* – chromosomes 1-20, X and Y; *R. rattus* and *A. niloticus* – chromosomes 1-18, X and Y.

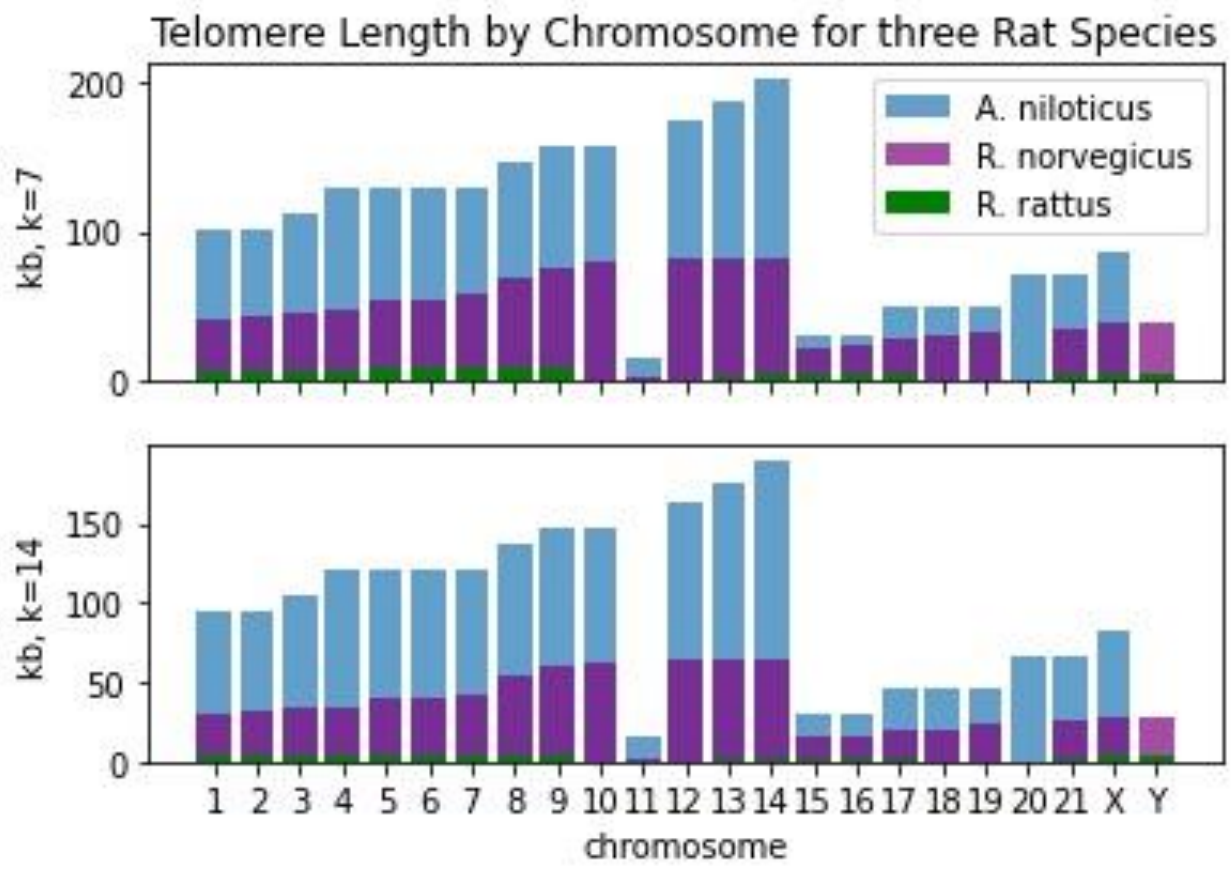

Figure 4:

Telomere length in kilobases by chromosome for three rat species, calculated from total base pairs contained in segments of at least k tandem repeats of TTAGGG. Not all species have the same number of chromosomes: *R. norvegicus* – chromosomes 1-20, X and Y; *R. rattus* and *A. niloticus* – chromosomes 1-18, X and Y.

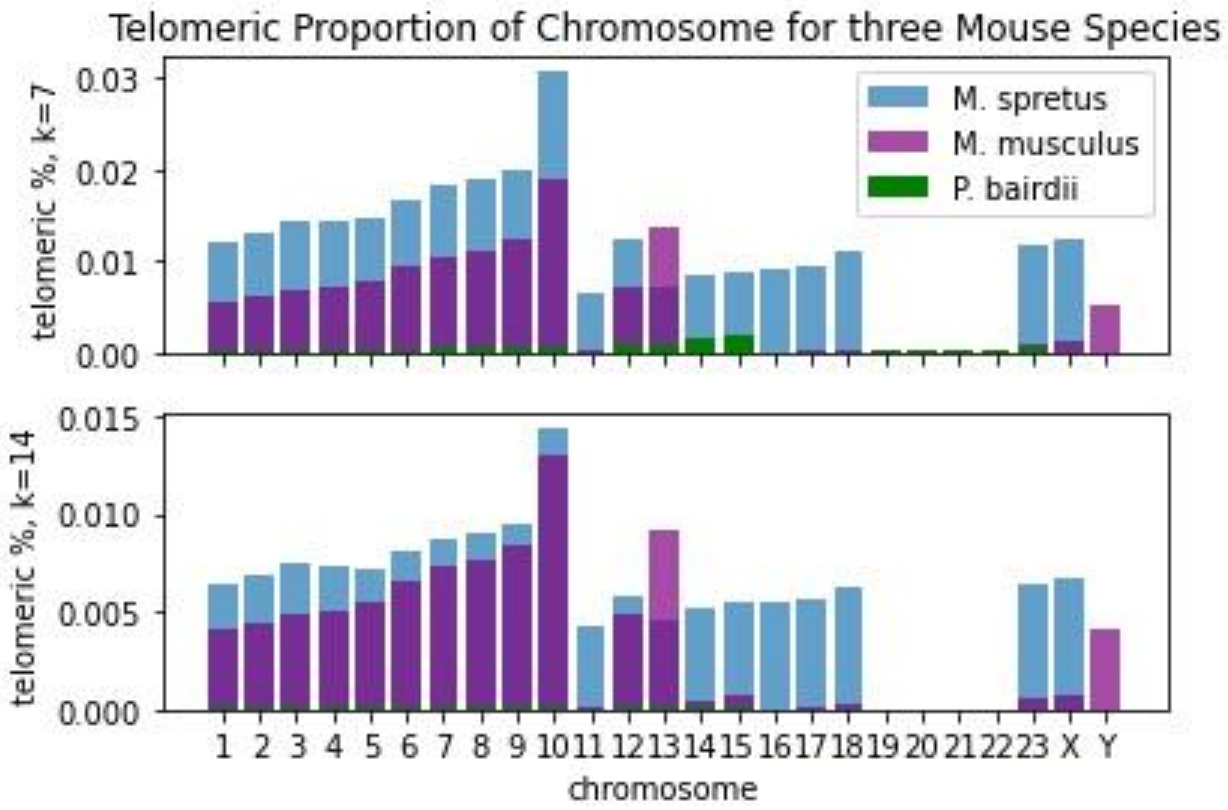

Figure 5:

Telomeric percentage of each chromosome for three mouse species, from total base pairs contained in segments of at least k tandem repeats of TTAGGG divided by length of the chromosome x 100%. Not all species have the same number of chromosomes (available in NCBI): *M. musculus* – chromosomes 1-19, X and Y; *M. spretus* – chromosomes 1-19, X; *P. bairdii* – chromosomes 1-18, X.

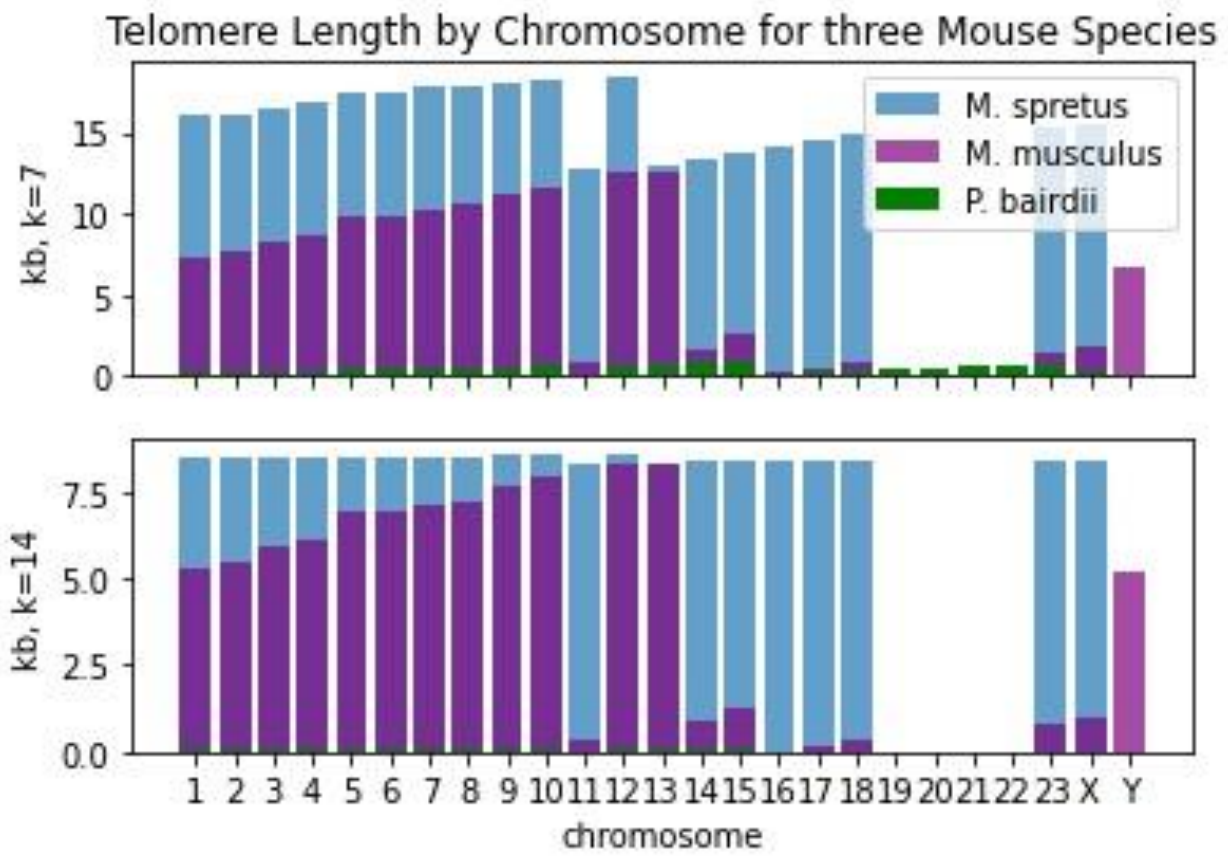

Figure 6:

Telomere length by chromosome for three mouse species, calculated from total base pairs contained in segments of at least k tandem repeats of TTAGGG. Not all species have the same number of chromosomes (available in NCBI): *M. musculus* – chromosomes 1-19, X and Y; *M. spretus* – chromosomes 1-19, X; *P. bairdii* – chromosomes 1-18, X.

Tables

Table 1:

Telomeric Proportion of Genome (TPG) for all species at k = 7. Left column: TPG in %. Columns 2 onward: TPG ratios for all pairs of species (row/column: for example, the upper right cell is the ratio of lab rat TPG to human TPG).

species	TPG (%)	R.n.	R.r.	A.n.	M.m.	M.s.	P.l.	P.m.b.	A.a.	O.c.	H.h.	H.s.
R.n.	0.040333	1	8.971	0.355	8.033	3.286	130.607	94.946	25.356	42.889	27.577	37.738
R.r.	0.004496	0.111	1	0.04	0.895	0.366	14.558	10.583	2.826	4.781	3.074	4.207
A.n.	0.113653	2.818	25.28	1	22.636	9.26	368.031	267.545	71.448	120.855	77.709	106.341
M.m.	0.005021	0.124	1.117	0.044	1	0.409	16.259	11.819	3.156	5.339	3.433	4.698
M.s.	0.012273	0.304	2.73	0.108	2.444	1	39.744	28.892	7.716	13.051	8.392	11.484
P.l.	0.000309	0.008	0.069	0.003	0.062	0.025	1	0.727	0.194	0.328	0.211	0.289
P.m.b.	0.000425	0.011	0.094	0.004	0.085	0.035	1.376	1	0.267	0.452	0.29	0.397
A.a.	0.001591	0.039	0.354	0.014	0.317	0.13	5.151	3.745	1	1.692	1.088	1.488
O.c.	0.00094	0.023	0.209	0.008	0.187	0.077	3.045	2.214	0.591	1	0.643	0.88
H.h.	0.001463	0.036	0.325	0.013	0.291	0.119	4.736	3.443	0.919	1.555	1	1.368
H.s.	0.001069	0.026	0.238	0.009	0.213	0.087	3.461	2.516	0.672	1.136	0.731	1
average	0.016507											

Table 2:

Total Telomeric Length (TTL) of genome for all species at k = 7. Left column: TTL in kb. Columns 2 onward: TTL ratios for all pairs of species (row/column).

species	TTL (kb)	R.n.	R.r.	A.n.	M.m.	M.s.	P.l.	P.m.b.	A.a.	O.c.	H.h.	H.s.
R.n.	1062.162	1	9.952	0.461	7.768	3.337	142.994	102.624	29.227	50.249	31.528	32.181
R.r.	106.728	0.1	1	0.046	0.781	0.335	14.368	10.312	2.937	5.049	3.168	3.234
A.n.	2302.656	2.168	21.575	1	16.84	7.235	309.997	222.479	63.361	108.934	68.348	69.765
M.m.	136.74	0.129	1.281	0.059	1	0.43	18.409	13.212	3.763	6.469	4.059	4.143
M.s.	318.252	0.3	2.982	0.138	2.327	1	42.845	30.749	8.757	15.056	9.446	9.642
P.l.	7.428	0.007	0.07	0.003	0.054	0.023	1	0.718	0.204	0.351	0.22	0.225
P.m.b.	10.35	0.01	0.097	0.004	0.076	0.033	1.393	1	0.285	0.49	0.307	0.314
A.a.	36.342	0.034	0.341	0.016	0.266	0.114	4.893	3.511	1	1.719	1.079	1.101
O.c.	21.138	0.02	0.198	0.009	0.155	0.066	2.846	2.042	0.582	1	0.627	0.64
H.h.	33.69	0.032	0.316	0.015	0.246	0.106	4.536	3.255	0.927	1.594	1	1.021
H.s.	33.006	0.031	0.309	0.014	0.241	0.104	4.443	3.189	0.908	1.561	0.98	1
average	369.8629											

Table 3:

TPG for all species at k = 14. At this threshold, *P. leucopus* and *H. hydrochaeris* have zero telomere content and are excluded from this dataset. Left column: TPG in %. Columns 2 onward: TPG ratios for all pairs of species (row/column).

species	TPG (%)	R.n.	R.r.	A.n.	M.m.	M.s.	P.m.b.	A.a.	O.c.	H.s.
R.n.	0.030303	1	13.843	0.283	8.868	4.636	627.817	80.802	454.089	74.026
R.r.	0.002189	0.072	1	0.02	0.641	0.335	45.352	5.837	32.802	5.347
A.n.	0.107213	3.538	48.978	1	31.374	16.401	2221.237	285.882	1606.582	261.906
M.m.	0.003417	0.113	1.561	0.032	1	0.523	70.799	9.112	51.208	8.348
M.s.	0.006537	0.216	2.986	0.061	1.913	1	135.43	17.43	97.954	15.969
P.m.b.	4.83E-05	0.002	0.022	0	0.014	0.007	1	0.129	0.723	0.118
A.a.	0.000375	0.012	0.171	0.003	0.11	0.057	7.77	1	5.62	0.916
O.c.	6.67E-05	0.002	0.03	0.001	0.02	0.01	1.383	0.178	1	0.163
H.s.	0.000409	0.014	0.187	0.004	0.12	0.063	8.481	1.092	6.134	1
average	0.016729									

Table 4:

TTL for all species at k = 14. At this threshold, *P. leucopus* and *H. hydrochaeris* have zero telomere content and are excluded from this dataset. Left column: TTL in kb. Columns 2 onward: TTL ratios for all pairs of species (row/column).

species	TTL (kb)	R.n.	R.r.	A.n.	M.m.	M.s.	P.m.b.	A.a.	O.c.	H.s.
R.n.	798.018	1	15.357	0.367	8.575	4.708	678.587	93.139	532.012	63.124
R.r.	51.966	0.065	1	0.024	0.558	0.307	44.189	6.065	34.644	4.111
A.n.	2172.174	2.722	41.8	1	23.34	12.815	1847.087	253.522	1448.116	171.822
M.m.	93.066	0.117	1.791	0.043	1	0.549	79.138	10.862	62.044	7.362
M.s.	169.5	0.212	3.262	0.078	1.821	1	144.133	19.783	113	13.408
P.m.b.	1.176	0.001	0.023	0.001	0.013	0.007	1	0.137	0.784	0.093
A.a.	8.568	0.011	0.165	0.004	0.092	0.051	7.286	1	5.712	0.678
O.c.	1.5	0.002	0.029	0.001	0.016	0.009	1.276	0.175	1	0.119
H.s.	12.642	0.016	0.243	0.006	0.136	0.075	10.75	1.475	8.428	1
average	367.6233									

Table 5:

Assemblies and accession IDs for each species in this study.

common name	species name	assembly	chromosomes	accession IDs
lab rat	Rattus norvegicus	mRatBN7.2	1-20, X, Y	NC_051336.1-NC_051357.1
black rat	Rattus rattus	Rrattus_CSIRO_v1	1-19, X, Y	NC_046154.1-NC_046173.1
African grass rat	Arvicanthis niloticus	mArvNil1.pat.X	1-21, X	NC_047658.1-NC_047679.1
lab mouse	Mus musculus	GRCm39	1-19, X, Y	NC_000067.7,NC_000068.8,NC_000069.7-NC_000085.7,NC_000086.8,NC_000087.8
Algerian mouse	Mus spretus	SPRET_Eij_v1	1-19, X	CM004094.1-CM004114.1
white-footed mouse	Peromyscus leucopus	UCI_PerLeu_2.1	1-23, X	NC_051063.1-NC_051086.1
prairie deer mouse	Peromyscus miniculatus bairdii	HU_Pman_2.1.3	1-23, X	NC_056008.1-NC_056030.1
European water vole	Arvicola amphibius	mArvAmp1.2	1-18, X	NC_052047.1,NC_052048.2,NC_052049.1,NC_052050.1,NC_052051.1,NC_052052.2,NC_052053.1,NC_052054.1,NC_052055.2,NC_052056.1,NC_052057.2,NC_052058.2,NC_052059.1,NC_052060.1,NC_052061.1,NC_052063.2,NC_052064.1,NC_052065.1
rabbit	Oryctolagus cuniculus	OryCun2.0	1-21, X	CM000790.1-CM000811.1
capybara	Hydrochoerus hydrochaeris	Hydrochoerus_hydrochaeris_HiC	1-33	CM027413.1-CM027445.1
human	Homo sapiens	GRCh38.p13	1-22, X, Y	NC_000001.11,NC_000002.12,NC_000003.12,NC_000004.12,NC_000005.10,NC_000006.12,NC_000007.14,NC_000008.11,NC_000009.12,NC_000010.11,NC_000011.10,NC_000012.12,NC_000013.11,NC_000014.9,NC_000015.10,NC_000016.10,NC_000017.11,NC_000018.10,NC_000019.10,NC_000020.11,NC_000021.9,NC_000022.11,NC_000023.11,NC_000024.10