Transposable elements activity reveals punctuated patterns of speciation in Mammals

Ricci Marco[1]* & Peona Valentina[1], Guichard Etienne[1], Taccioli Cristian[2], Boattini Alessio[1]
[1]University of Bologna, Department of Biological, Geological and Environmental Sciences, Italy
[2]University of Padova, Department of Animal Medicine, Production an Health, Italy
*corresponding author: marco.ricci19@unibo.it

Abstract

Transposable elements (TEs) play an essential role in shaping eukaryotic genomes and in organism diversification. It has been hypothesized that bursts of TEs activity may correspond to punctuated events of speciation (CArrier SubPopulation, Epi-Transposon, TE-Thrust hypotheses), thus it is expected that highly differentiated taxa might bear highly active TEs in their genomes. Two new parameters designed to measure the taxa adaptive radiation and the magnitude of TE activity were created: the Relative Rate of Speciation (RRS) and the Density of Insertion (DI). Furthermore, we defined as “hot” and “cold” those genomes with high and low DI respectively. The correlation between RRS and DI, which we called "Cold Genome" hypothesis, was tested in Mammalian families and superorders. Since ages of TEs on a large scale can be approximated by calculating their distance from the respective consensus sequences, we subsetted TEs in different classes in order to study the evolution of genomes at different time scales. We considered “recent” those TEs with < 1% divergence, whereas we called “less recent” TEs with < 5% divergence. Comparing the TEs activity in 16 pairs of species belonging to different mammalian families and the average TEs activity of sampled species from the four superorders of Placentalia, we showed that taxa with positive RRS correlate with “hot genomes”, whereas taxa with negative RRS correlate with “cold genomes”. Specifically, the density of recent insertions correspond to recent macroevolutionary events, while the density of less recent insertions coincide with older events. These results are fully coherent with our “Cold Genome” hypothesis. In addition, our study supports, in both phases of radiation and stasis, the “Punctuated Equilibria”

theory in mammals.

Introduction

1. Gradualism and Punctuated Equilibria

The debate between phyletic gradualism and punctuated equilibria (PE) in evolutionary biology is still intense[1-3]. Phyletic gradualism was embraced by the Modern Synthesis as the evolutionary dynamics for speciation implying a gradual accumulation of mutation throughout time, until genetic incompatibilities break up the gene flow. The theory of punctuated equilibria proposes, instead, that "[…] evolution is concentrated in very rapid events of speciation (geologically instantaneous, even if tolerably continuous in ecological time). Most species, during their geological history, either do not change in any appreciable way, or else they fluctuate mildly in morphology, with no apparent direction"[4]. Evidences in favor of punctuated equilibria come from diverse fields such as paleontology[4-6], phylogenesis[7,8] and experimental evolution[9]. At the same time, gradualism gained evidences when studying various living models[11,12].

2. The evolutive power of TEs

Transposable Elements (TEs) are far from being junk DNA[13]. In the last years they have been continuously linked to essential cellular activities such as the telomeres repairing[13-15], rewiring of transcriptional networks[16,17], regulation of gene expression[18,19], ectopic recombination and chromosomal rearrangements[20]. TEs are key contributors to evolution and play/played a fundamental role in biological processes of utmost importance[13,21-27], like the insurgence of the V(D)J system of acquired immunity[22,28,29] , placenta development[30], embryogenesis[31,32] and neurogenesis[33-36]. Mobile elements make genomes fluid, dynamic[37] and organisms evolvable[38,39].

Given their huge impact on shaping genomes, Transposable Elements are thought to contribute to the formation of reproductive barriers facilitating speciation[32,40-43]. Some authors proposed to correlate high TEs activity with organismal differentiation[44-46]. Moreover, the environment and its influence on the epigenetic structure of the cell seems to modulate the mobilization of TEs. The disruption of the epigenome potentially leads to bursts of activity and to a rapid accumulation of genomic variability necessary for phenotypic innovation and speciation (Epi-Transposon and TE-thrust hypotheses)[44,45]. Furthermore, the diversification of TE families is likely to coincide with events of genetic drift (CArrier SubPopulation hypothesis)[46]. Interestingly, these observations[44-46] can explain both gradualistic and punctuated evolution on the premise of TEs content in genomes. The debate between gradualism and punctuated equilibria is still open but we believe that TEs evolutionary studies can shed some light on it.

3. Hot and cold genomes

Organisms owe their ability to diversificate to their genomic plasticity and the activity of TEs can substantially contribute to it[20,21]. Hence, we expect that a positive relationship exists between TEs activity and the extant biodiversity. For example, within Mammals, the order Monotremata is the most ancient and the poorest in living species (Figure 1); accordingly, the platypus, that belongs to this group, has a genome that harbors the lowest number of recently mobilized elements[46]. Is it possible that taxa with low rates of speciation are associated with genomes with inactive TEs? Starting from the observations of specific cases, we widen the perspective to a general evolutionary explanation that we called the "Cold Genome" hypothesis. According to this hypothesis, genomes with highly active TEs ("hot genomes") belong to taxa with high rates of speciation, whereas genomes with inactive TEs ("cold genomes") belong to taxa with low rates of speciation (Figure 2A). We in-

vestigated the TEs activity using the Density of Insertion (DI) of elements and (as previously proposed in literature[44]) the number of TE families at different divergences from consensus. Furthermore, we introduced the concept of Relative Rate of Speciation (RRS) in order to establish the magnitude of speciation rates within a given taxa (Figure 2B).

Viable elements act like an evolutionary driving force (leading to punctuated events/bursts of insertions/"hot" genomes) but cells have a plethora of molecular mechanisms that modulate the expression of TEs. Molecular mechanisms seem to intervene in temporally discrete periods to regulate the activity of TEs[45] - for example young LINE-1 elements are repressed via methylation while ancient TEs are repressed by the KRAB/KAP1 system[47] - potentially leading to a paucity of innovative elements and a macroevolutionary genomic stasis ("cold" genomes).

In this paper, we investigated the mechanisms of speciation in Mammals, showing how the gradualistic model does not seem to be able to fully describe the evolution of the extant biodiversity, whereas our new evolutionary hypothesis, that includes PE theory and the genomic impact of transposable elements, might better explain evolutionary dynamics.

Results

4. Clade age is not related to species richness in Mammalia

From a neodarwinian point of view, species continuously accumulate mutations that would eventually lead to differentiation and speciation; therefore, older taxonomical groups should have had more time to accumulate biodiversity, leading to an overabundance of species in comparison to younger groups[12]. We tested the phyletic gradualism model in Mammalia by investigating the relationship between clade age and species richness of the

152 mammalian families, using both a linear regression model (*lm* function in R, stats package[47]) and a non-parametric correlation test (*cor.test* function in R[47]). We retrieved the number of species for all the 152 mammalian families listed in the last mammalian phylogeny[49] from Catalogue of Life[50]. The crown age of the mammalian families was estimated from their timed phylogenetic tree[49]. The calculated regression coefficient is slightly negative (-0.1104) and the R^2 is very low (0.00041, P-value 0.8039), hence there is no statistically significant association between the two variables (Figure S1). There is no significant correlation between clade age and species richness either (rho 0.01815343, P-value 0.8243). This model does not seem to describe mammalian evolution accurately, as their differentiation pattern seems to behave in a more complex way.

5. TEs activity and speciation correlate in the whole Mammalia class

TEs are powerful facilitators of evolution[20-24,44,45] and they are tightly associated to the evolutionary history of mammals[16,30,31]. The data about TE families and TE insertions in the genomes of the species considered were analyzed by Jurka and colleagues[46].

For each TE family, a consensus sequence was produced[51], representing the reference sequence for that element. High divergence of a sequence from its consensus approximates, on a large scale, the long time it had to accumulate mutations, whereas lower divergence represents a more recent and less pronounced molecular differentiation. The mobile elements diverging less than 1% from their consensus sequences and their respective TE families were pooled in the "1% dataset". This dataset represents the most recently mobilized elements, whereas the ones that diverge less than 5% were included in the "5% dataset", i.e. the list of both recent and more ancient insertions. For better evaluation of the activity of mobile elements in the considered genomes/species, we propose a new para-

meter herein called Density of Insertion (DI). DI is calculated according to the formula: DI = NI / GS, where NI is the total Number of Insertions (of elements contained in the 1% or 5% datasets) and GS is the Genome Size in gigabases. Accordingly, DI is measured in insertions for gigabase (ins/Gb). We call "1% DI" the parameter calculated using the "1% dataset" and "5% DI" the one that used the "5% dataset". All the parameters measuring TEs activity were averaged between species belonging to the same taxonomical family, which allowed to perform analyses on a larger scale.

In order to test if and how TEs activity reflects mammalian speciation pattern, we calculated the Rate of Speciation (RS) with the formula: RS = NS / CA, where NS is the total number of species for the analyzed taxonomical family, whereas CA is the Crown Age of the same taxon[49].

In Figure 3, we show the TEs activity measured for all the families. These families are arranged in order of RS. There is an increasing trend for all the parameters from left (low Rate of Speciation) to right (high Rate of Speciation). This possible association was tested through non-parametric statistical correlation (using R[47] function *cor.test* with the Spearman method) and by generating a linear regression model (*lm* function in R[47], stats package) (Table S2). All the parameters show significant correlation with the Rate of Speciation in the whole Mammalia class (Table S2). In order of significance, the parameters with better descriptive power are: 5% DI (P-value < 0.005), 1% dataset families (P-value < 0.005), 5% dataset families (P-value < 0.05) and 1% DI (P-value < 0.05). Linear regression models for all the parameters in function of the Rate of Speciation were estimated (Table S3). All models show positive angular coefficients and have significant P-values. These results indicate that TEs activity seems to be tightly associated with speciation events.

6. Families with higher Relative Rate of Speciation show

higher TE activity

Mammalian phylogeny seems to support an evolutionary framework in which short bursts of diversification are alternated with longer periods of relative stasis, similarly to what is stated in the punctuated equilibria theory[4-6].

In order to validate this observation, we tested two key factors: the activity of Transposable Elements (using the same parameters described above), and the newly introduced Relative Rate of Speciation (RRS). In fact, RS does not factor in when speciation occurred intra-taxonomically, intrinsically including a bias that can offset the results of evolutionary researches. The RRS, instead, is a binomial parameter, also based on the age of the clade of interest and its species richness, that, given a pair of taxa, identifies which one has the highest speciation activity. Briefly, the taxon that shows a higher number of species and, at the same time, is younger has a positive (+) Relative Rate of Speciation and a putative "hot" genome; consequently the other taxon has a negative (-) RRS and a putative "cold" genome. All comparisons were performed between species belonging to the same taxonomical order. In this way the compared species share a common history until the origin of their respective orders.

The RRS attribution can be represented by the logical formulae:

$RRS_1(+) \leftarrow NS_1 > NS_2 \wedge CA_1 < CA_2$

$RRS_1(-) \leftarrow NS_1 < NS_2 \wedge CA_1 > CA_2$

If neither of these assertions is true, the RRS is non-applicable (NA), and the pair of taxa cannot be compared because of the impossibility to distinguish the age of their intra-taxonomical differentiation. This enables the exclusion of non-informative (or misleading) information while still retaining the ability to calculate a reliable proxy of evolvability with only two basic parameters.

As an explanatory example, we can describe in detail some comparisons in the order

primates (Figure 2B). Galago, that includes 19 extant species, is a monkey of the family Galagidae. This family is more ancient than Cercopitecidae, whereas the number of species in Cercopitecidae is higher than Galagidae. In this particular case, Galagidae has RRS(+) compared to Cercopitecidae (Figure 2AB). When comparing Galagidae and Tarsidae families, the situation was the opposite: Tarsidae is more ancient and poorer in species than Galagidae, so the latter has RRS(+) (Figure 2B). It is not always possible to determine RRS for all the potential pairs of families/species. For instance, we could not compare the families Hominidae and Cercopithecidae. In particular, Hominidae seems to have a lower rate of speciation, with only 7 living species when compared the Cercopithecidae's that includes 159 species; at the same time the family Cercopithecidae is 8 Mya older, which means that this taxon could have accumulated more species in a larger amount of time.

In total, we tested our hypothesis on 16 families, represented by 19 species, that encompass six mammalian orders (Table 1).

The four parameters of TEs activity, used as proxies for evolvability and differentiability as aforementioned, were measured. The levels of "1% DI" dataset in all the observed pairs are shown in Figure 4A. The correspondence between putative "hot"/"cold" genomes, based on the RRS, and "hotness" and "coldness" of the same genomes predicted by the four TEs activity parameters was tested with a paired Wilcoxon Signed Rank Test (using the *wilcox.test* function in R): P-values are significant with the exception of "5% DI" dataset (Table 3). In order of significance (and descriptive efficiency), the parameters that better explain the relationship between "hot"/"cold" genomes and TEs content are (P-values in parentheses): "1% DI" dataset (0.0013), "5% TE" Families (0.0063), "1% TE" Families (0.0075) and "5% DI" dataset (0.0739). The descriptive efficiency of the chosen parameters is also presented in Figure 4B. When there is no association between the better descriptive parameter ("1% DI") and RRS, the other parameters show no association either.

Within the "1% DI" dataset, 14 out of 16 pairs follow the expected trend of association between DI values and RRS. Among those, 11 pairs show a difference in DI of at least one order of magnitude, up to almost 180-fold higher in the pair *Macaca mulatta - Tarsius syrichta*. Despite the two exceptions, *Microcebus murinus - Callithrix jacchus* and *Otolemur garnettii - Callithrix jacchus*, the statistical support clearly suggests that, in Mammals, TEs are associated with adaptive radiation.

It is worth noticing that *Canis lupus* is the only species that, despite the lower total number of mobile elements insertions (lower DI: 194 ins/GB), has more specific TE families (4) than its paired species *Felis catus* (higher DI: 1446 ins/GB, 3 specific TE families). The other case in which the "1% DI" and the number of 1% dataset TE families show discordance is the pair *Tarsius syrichta - Otolemur garnettii*. *Otolemur garnettii* has in fact more new integrants (DI: 51 Ins/Gb) compared to *Tarsius syrichta* (DI: 6 Ins/Gb) but the same number of TE families (2 each). In both these cases, the greater diversity in TE families (or absence thereof) does not reflect the relative hotness of the species genome, which is better described by the impact of the elements on their genome. From the example presented and the statistical tests, DI is a more sensible parameter than the abundance of TEs families, since the diversity in active families is not always a proxy of the activity of mobile elements in the genome.

The parameters measuring TEs activity confirmed to be accurate even at an intra-family level (as shown in the case of Muridae). Indeed, keeping as constant the taxa with negative RRS and interchanging different genera of the same family as a comparison point, we observed variations in the parameters values, but a statistically similar relevance of their association with the RRS (Figure 4B).

The results obtained point to state that the activity of TEs does not vary randomly within the mammalian phylogeny and that the RRS shows strong association with this activity. Therefore, RRS allows prediction of the relative level of TE mobility between two taxa,

which in turn is highly related to their ability to differentiate and speciate.

7. Ancient TE bursts correlate with the ancient history of Placentalia

Once proven that DI is the most descriptive parameter for intra-order activity of TEs, we tested our hypothesis at a higher taxonomic level.

The combined groups consist of 22 species belonging to "hot" superorders and 5 belonging to "cold" superorders. For both 1% and 5% DI averaged datasets, Xenarthra (X) and Afrotheria (A) have a mean DI more than tree fold lower than Laurasiatheria (L) and Euarchontoglires (E). Specifically for the 5% dataset, the pairs A - E and X - E show a ratio of about 1:5, while for the pairs A - L and X - L the ratio is about of 1:8. Minor differences can be observed for the average density of insertions within the "1% DI" dataset. In increasing order of ratio we have X - E (1:3), E - L (1:4), A - E (1:7), A - L (1:10). The comparison between (X + A) and (E + L) at 1% of divergence from the consensus is non-significant with a P-value of 0.0973.

Bar plots in Figure 4C show that, in both comparisons, standard errors of average DI do not overlap. The Wilcoxon test, performed for the average "5% DI" parameter, showed significant difference between the two groups (P-value 0.0394). Although the same tendency is clearly visible for the average "1% DI" parameter, the Wilcoxon give marginally non-significant results: the putative colder taxa are well differentiated in absolute values but the standard errors of the pairs X - A and X - E overlap one another only slightly.

It must be noted that, given the biological paucity of species in the groups Xenarthra and Afrotheria, their 1% and 5% datasets are very small and heterogeneous (i.e. low statistical power). For this reason, it is hard to statistically compare them to the larger datasets of Euarchontoglires and Laurasiatheria.

The discrepancy between the results of the two datasets may be interpreted from an evolutionary point of view. The Density of Insertion at 5% is the least accurate parameter to study recent events (Figure 4B and Table 3) but it works efficiently in the study of older macroevolutive events such as the origin of the four superorders of Placentalia (Figure 4C). The divergence of the elements from their consensus tends, in average, to reflect their age and thus the different datasets describe different periods of time and related taxonation events.

Conclusions

Transposable Elements are a major source of genomic variability and have more than once impacted evolution through the rise of key molecular processes[13-33,36]. Even though they seem to be an important factor for the adaptive radiations, the genomic signal they left in the extant biodiversity is exposed, as many biological phenomena, to alteration and noise throughout time. In evolutionary time-scales, their activity is modulated producing alternations of insertional bursts and silencings, which is consistent with speciation patterns explained by the punctuated equilibria theory. Phenotypic differentiation and adaptive radiation would thus macroscopically reflect TEs activity molecular dynamics. Furthermore, TEs seem to positively influence speciation, as is shown by the results of our study: a high differentiation rate is strongly associated with an increased molecular activity of the mobilome. This is apparent for both general and relative rates of speciation (RS and RRS) and reflected by all the parameters used as proxies for TEs activity and the states of "hot" or "cold" genomes, with Density of Insertion being the most descriptive.

When silencing mechanisms would progressively inhibit TEs activity (state of "cold" genome), their lack of contribution to molecular differentiation (negative RRS) seems to lead to the relatively static phase postulated by punctuated equilibria.

The "Cold Genome" hypothesis thus supports the punctuated equilibria theory in both the punctuated differentiation bursts and stasis periods.

Furthermore, we showed that TEs insertions describe the insurgence of clades accordingly to their estimated age: "less recent" TE bursts are a proxy for older taxonation events (mammalian superorders), "recent" ones for late events (mammalian families).

Whether TEs mobilization and accumulation of new insertions is cause or effect of adaptive radiation and speciation remains open for debate, although the results presented (and the intrinsic characteristics of the mobilome's activity) seem to suggest the former of the two to be the more likely hypothesis.

References

1. Pennell, M., Harmon, L. & Uyeda, J. Is there room for punctuated equilibrium in macroevolution? *Trends in Ecology & Evolution* 29, 23-32 (2014).

2. Lieberman, B. & Eldredge, N. What is punctuated equilibrium? What is macroevolution? A response to Pennell et al. *Trends in Ecology & Evolution* 29, 185-186 (2014).

3. Casane, D. & Laurenti, P. Why coelacanths are not 'living fossils'. *Bioessays* 35, 332-338 (2013).

4. Gould, S. & Eldredge, N. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3, 115-151 (1977).

5. Eldredge, N. & Gould, S. Punctuated equilibrium: an alternative to phyletic gradualism. *Models in Paleobiology* 82-115 (1972).

6. Gould, S. *The structure of evolutionary theory*. (Belknap Press of Harvard University Press, 2002).

7. Pagel, M., Venditti, C. & Meade, A. Large Punctuational Contribution of Speciation to Evolutionary Divergence at the Molecular Level. *Science* 314, 119-121 (2006).

8. Mattila, T. & Bokma, F. Extant mammal body masses suggest punctuated equilibrium. *Proceedings of the Royal Society B: Biological Sciences* 275, 2195-2199 (2008).

9. Lenski, R. & Travisano, M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences* 91, 6808-6814 (1994).

11. McPeek, M. & Brown, J. Clade Age and Not Diversification Rate Explains Species Richness among Animal Taxa. *The American Naturalist* 169, E97-E106 (2007).

12. Stadler, T., Rabosky, D., Ricklefs, R. & Bokma, F. On Age and Species Richness of Higher Taxa. *The American Naturalist* 184, 447-455 (2014).

13. Biemont, C. A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution. *Genetics* 186, 1085-1093 (2010).

14. Farkash, E. & Prak, E. DNA Damage and L1 Retrotransposition. *Journal of Biomedicine and Biotechnology* 2006, 1-8 (2006).

15. Muotri, A., Marchetto, M., Coufal, N. & Gage, F. The necessary junk: new functions for transposable elements. *Human Molecular Genetics* 16, R159-R167 (2007).

16. Lynch, V., Leclerc, R., May, G. & Wagner, G. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics* 43, 1154-1159 (2011).

17. Kunarso, G. et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics* 42, 631-634 (2010).

18. Faulkner, G. et al. The regulated retrotransposon transcriptome of mammalian cells. *Nature Genetics* 41, 563-571 (2009).

19. Slotkin, R. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics* 8, 272-285 (2007).

20. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics* 9, 397-405 (2008).

21. Fedoroff, N. Transposable Elements, Epigenetics, and Genome Evolution. *Science* 338, 758-767 (2012).

22. Kapitonov, V. & Jurka, J. RAG1 Core and V(D)J Recombination Signal Sequences Were Derived from Transib Transposons. *PLoS Biology* 3, e181 (2005).

23. Feschotte, C. & Pritham, E. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* 41, 331-368 (2007).

24. Emera, D. & Wagner, G. Transposable element recruitments in the mammalian placenta: impacts and mechanisms. *Briefings in Functional Genomics* 11, 267-276 (2012).

25. Piskurek, O. & Jackson, D. Transposable Elements: From DNA Parasites to Architects of Metazoan Evolution. *Genes* 3, 409-422 (2012).

26. Richardson, S. et al. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiology Spectrum* 3, (2015).

27. Carmi, S., Church, G. & Levanon, E. Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. *Nature Communications* 2, 519 (2011).

28. Koonin, E. & Krupovic, M. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet* 16, 184-192 (2014).

29. Huang, S. et al. Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* 166, 102-114 (2016).

30. Lavialle, C. et al. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20120507-20120507 (2013).

31. Gerdes, P., Richardson, S., Mager, D. & Faulkner, G. Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol* 17, (2016).

32. Friedli, M. & Trono, D. The Developmental Control of Transposable Elements and the Evolution of Higher Species. *Annual Review of Cell and Developmental Biology* 31, 429-451 (2015).

33. Notwell, J., Chung, T., Heavner, W. & Bejerano, G. A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nature Communications* 6, 6644 (2015).

34. Levin, H. & Moran, J. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* 12, 615-627 (2011).

35. Singer, T., McConnell, M., Marchetto, M., Coufal, N. & Gage, F. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends in Neurosciences* 33, 345-354 (2010).

36. Erwin, J., Marchetto, M. & Gage, F. Mobile DNA elements in the generation of diversity and complexity in the brain. *Nature Reviews Neuroscience* 15, 497-506 (2014).

37. Huff, J., Zilberman, D. & Roy, S. Mechanism for DNA transposons to generate introns on genomic scales. *Nature* 538, 533-536 (2016).

38. Schrader, L. et al. Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nature Communications* 5, 5495 (2014).

39. Hof, A. et al. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102-105 (2016).

40. Belyayev, A. Bursts of transposable elements as an evolutionary driving force. *Journal of Evolutionary Biology* 27, 2573-2584 (2014).

41. Senerchia, N. et al. Differential introgression and reorganization of retrotransposons in hybrid zones between wild wheats. *Molecular Ecology* 25, 2518-2528 (2016).

42. Oliver, K., McComb, J. & Greene, W. Transposable Elements: Powerful Contributors to Angiosperm Evolution and Diversity. *Genome Biology and Evolution* 5, 1886-1901 (2013).

43. Feiner, N. Accumulation of transposable elements in Hoxgene clusters during adaptive radiation of *Anolis lizards*. *Proceedings of the Royal Society B: Biological Sciences* 283, 20161555 (2016).

44. Oliver, K. & Greene, W. Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TE-Thrust hypothesis. *Ecol Evol* 2, 2912-2933 (2012).

45. Zeh, D., Zeh, J. & Ishida, Y. Transposable elements and an epigenetic basis for punctuated equilibria. *BioEssays* 31, 715-726 (2009).

46. Jurka, J., Bao, W. & Kojima, K. Families of transposable elements, population structure and the origin of species. *Biology Direct* 6, 44 (2011).

47. Castro-Diaz, N., Friedli, M. & Trono, D. Drawing a fine line on endogenous retroelement activity. *Mobile Genetic Elements* 5, 1-6 (2015).

48. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

49. Meredith, R. et al. Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science* 334, 521-524 (2011).

50. Catalogue of Life. *Catalogueoflife.com* (2016). at <http://www.catalogueoflife.com>

51. Bao, W., Kojima, K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* 6, (2015).

Figures and Tables

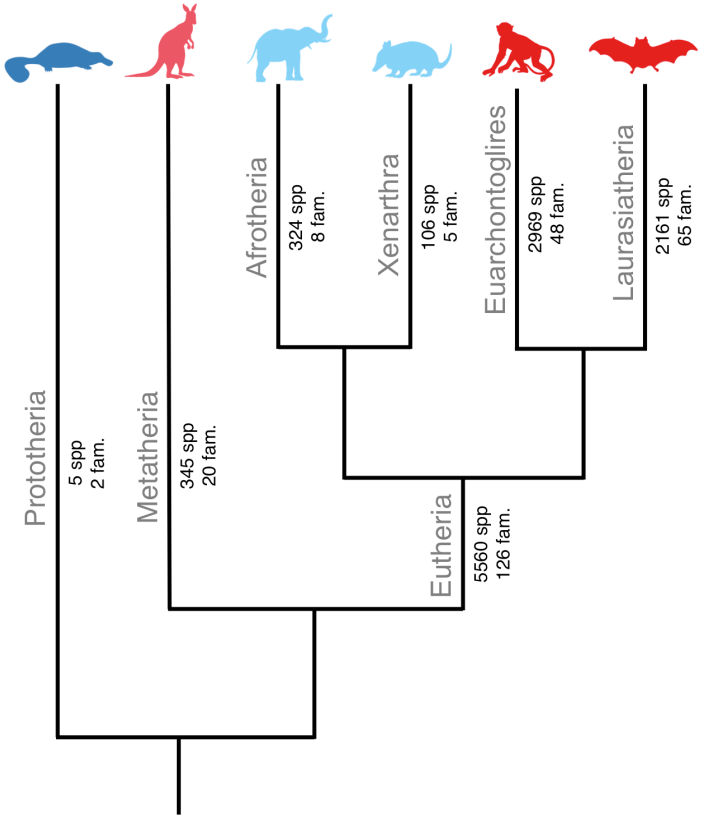


Figure 1. Tree of Mammals. Species abundance and phylogenetic relationships of the main mammalian clades. Animal icons made by Freepik from www.flaticon.com

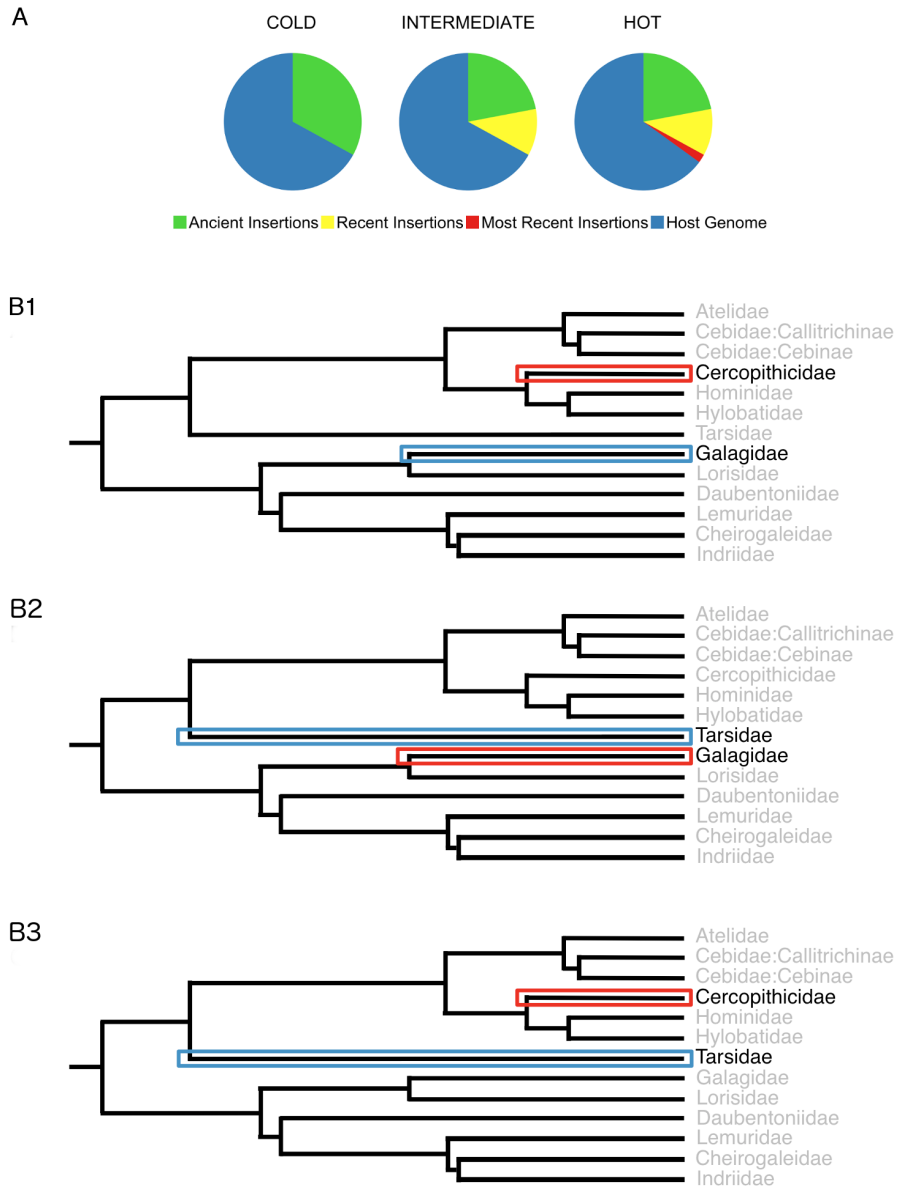


Figure 2. A) Modelization of the Cold Genome hypothesis. The cold genome presents only non-viable TEs. The intermediate genome model contains a fraction of TEs recently mobilized (divergent less than 5% from their consensus sequence). The hot genome has also a fraction of TEs more recently mobilized (divergent less than 1% from their consensus sequence).
B) Exemplified Relative Rate of Speciation in families of the order Primates. B1) Galagidae is older and poorer in species than Cercopithecidae so it has a lower RRS. Since our parameter is relative, in B2) Galagidae is younger and richer in species than Tarsidae, so in this case Galagidae has a higher RRS than Tarsidae. B3) Cercophitecidae is younger and richer in species than Tarsidae so it has a higher RRS than Tarsidae.

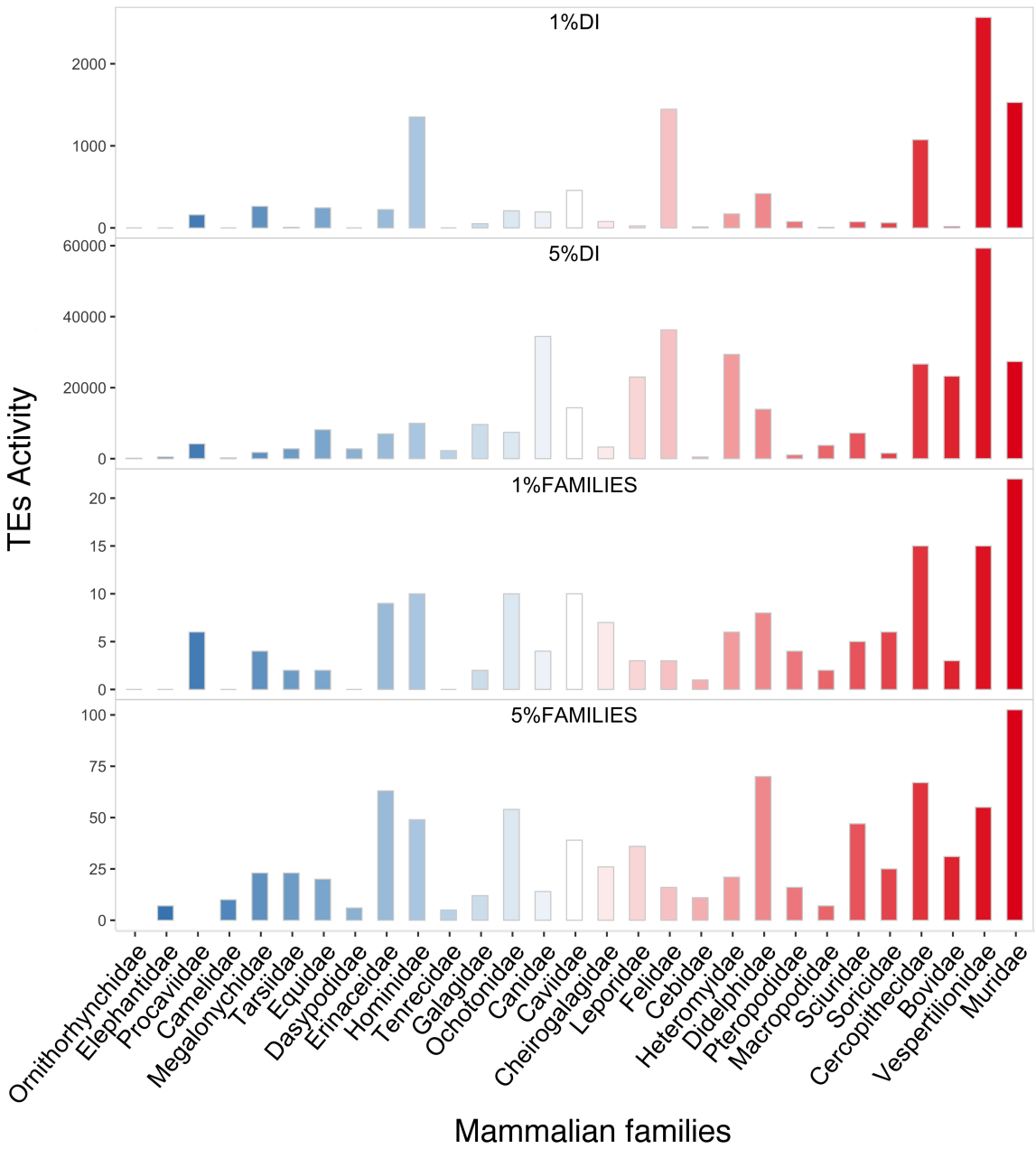
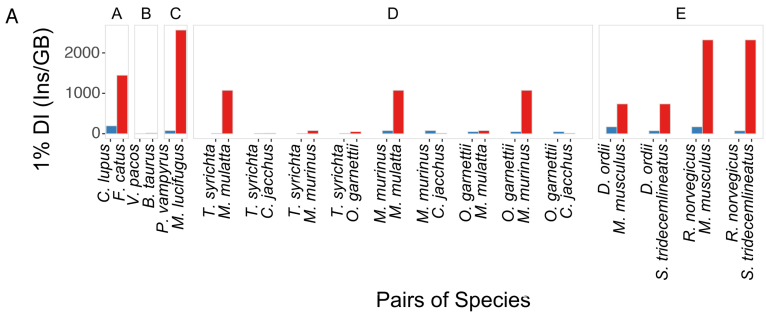


Figure 3. Relationship between the rate of speciation (RS) and TE activity in 29 mammalian families. The families are arranged in crescent order of RS.



B

Low RRS	High RRS	DI 1%	Families 5%	Families 1%	DI 5%
Canis lupus	*Felis catus*	✓	✓	✗	✓
Lama pacos	*Bos taurus*	✓	✓	✓	✓
Pteropus vampyrus	*Myotis lucifugus*	✓	✓	✓	✓
Tarsius syrichta	*Macaca mulatta*	✓	✓	✓	✓
Tarsius syrichta	*Callithrix jacchus*	✓	✗	✗	✗
Tarsius syrichta	*Microcebus murinus*	✓	✓	✓	✓
Tarsius syrichta	*Otolemur garnettii*	✓	✗	✗	✓
Microcebus murinus	*Macaca mulatta*	✓	✓	✓	✓
Microcebus murinus	*Callithrix jacchus*	✗	✗	✗	✗
Otolemur garnetti	*Macaca mulatta*	✓	✓	✓	✓
Otolemur garnettii	*Microcebus murinus*	✓	✓	✓	✗
Otolemur garnettii	*Callithrix jacchus*	✗	✗	✗	✗
Dipodomys ordii	*Mus musculus*	✓	✓	✓	✗
Dipodomys ordii	*Rattus norvegicus*	✓	✓	✓	✗
Spermophilus tridecemlineatus	*Mus musculus*	✓	✓	✓	✓
Spermophilus tridecemlineatus	*Rattus norvegicus*	✓	✓	✓	✓

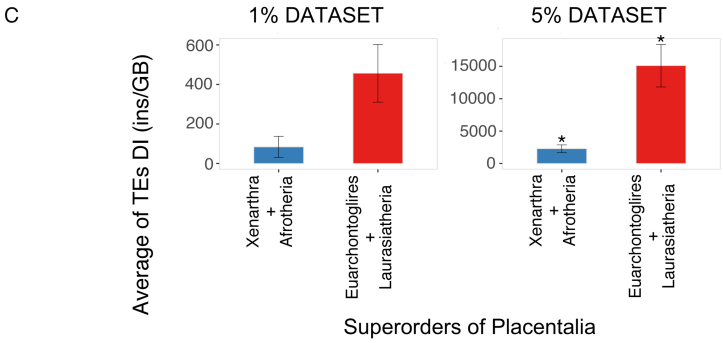


Figure 4. A) Comparison of the DI 1% in the 19 pairs of Mammals. Blue bars: putative cold genomes; Red Bars: putative hot genomes. The barplot is divided per order: A) Carnivora, B) Cetartiodactyla, C) Chiroptera, D) Primates, E) Rodentia.
B) Efficiency of the four parameters used. Green ticks and red crosses indicate, respectively, the pairs of species where the model fits or not.
C) Comparison of the DI 1% and DI 5% in the 4 superorders of Placentalia. Blue bars: putative cold genomes; Red Bars: putative hot genomes. * pvalue < 0.05

Table 1. The 19 pairs of selected species. For each species: the order, family, family age and family number of species are shown.

	LOWER RRS				HIGHER RRS			
ORDER	**FAMILY**	**N SPECIES**	**AGE (MY)**	**SPECIES**	**FAMILY**	**N SPECIES**	**AGE (MY)**	**SPECIES**
Carnivora	Canidae	35	47	*Canis lupus* Linnaeus, 1758	Felidae	42	30	*Felis catus* Linnaeus, 1758
Cetartiodactyla	Camelidae	5	62.5	*Lama pacos* Linnaeus, 1758	Bovidae	144	15	*Bos taurus* Linnaeus, 1758
Chiroptera	Pteropodidae	195	62	*Pteropus vampyrus* Linnaeus, 1758	Vespertilionidae	456	45	*Myotis lucifugus* Le Conte, 1831
Primates	Tarsiidae	11	64	*Tarsius syrichta* Linnaeus, 1758	Cercopithecidae	159	20	*Macaca mulatta* Zimmermann, 1780
				Tarsius syrichta Linnaeus, 1758	Cebidae	29	9.5	*Callithrix jacchus* Linnaeus, 1758
				Tarsius syrichta Linnaeus, 1758	Cheirogaleidae	34	36	*Microcebus murinus* J. F. Miller, 1777
				Tarsius syrichta Linnaeus, 1758	Galagidae	19	30	*Otolemur garnettii* Ogilby, 1838
	Cheirogaleidae	34	36	*Microcebus murinus* J. F. Miller, 1777	Cercopithecidae	159	20	*Macaca mulatta* Zimmermann, 1780
				Microcebus murinus J. F. Miller, 1777	Cebidae	29	9.5	*Callithrix jacchus* Linnaeus, 1758
	Galagidae	19	30	*Otolemur garnettii* Ogilby, 1838	Cercopithecidae	159	20	*Macaca mulatta* Zimmermann, 1780
				Otolemur garnettii Ogilby, 1838	Cheirogaleidae	34	36	*Microcebus murinus* J. F. Miller, 1777
				Otolemur garnettii Ogilby, 1838	Cebidae	29	9.5	*Callithrix jacchus* Linnaeus, 1758
Rodentia	Heteromyidae	65	35	*Dipodomys ordii* Woodhouse, 1853	Muridae	755	21	*Mus musculus* Linnaeus, 1758
				Dipodomys ordii Woodhouse, 1853				*Rattus norvegicus* Berkenhout, 1769
	Sciuridae	286	52	*Spermophilus tridecemlineatus* Mitchill, 1821				*Mus musculus* Linnaeus, 1758
				Spermophilus tridecemlineatus Mitchill, 1821				*Rattus norvegicus* Berkenhout, 1769

Table 2. The four superorder of Mammals with the list of their species used in this study. The number of species and age for each superorder are indicated.

LOWER RRS			HIGHER RRS		
SUPERORDER	**N SPECIES**	**AGE (MY)**	**SUPERORDER**	**N SPECIES**	**AGE (MY)**
Afrotheria	37	97	Euarchontoglires	1943	90
Echinops telfair			*Callithrix jacchus*		
Loxodonta africana			*Cavia porcellus*		
Procavia capensis			*Dipodomys ordii*		
Xenarthra	35	97	*Homo sapiens*		
Choloepus ofmanni			*Macaca mulatta*		
Dasypus novemcinctus			*Mus musculus*		
			Otolemur garnettii		
			Ochotona princeps		
			Orictolagus cuniculus		
			Microcebus murinus		
			Pan troglodytes		
			Rattus norvegicus		
			Spermophilus tridecemlineatus		
			Tarsius syrichta		
			Laurasiatheria	1113	90
			Bos taurus		
			Canis lupus familiaris		
			Equus caballus		
			Erinaceus europaeus		
			Felis catus		
			Lama pacos		
			Myotis lucifugus		
			Pteropus vampyrus		
			Sorex araneus		

Table 3. Results from the Wilcoxon rank sum test and statistical significance for each parameter used. N Higher RRS and N Lower RRS: abundance of the sets of values compared belonging to species with high or low RRS.

	N Higher RRS	N Lower RRS	P-value
DI 1%	16	16	0.001312
DI 5%	16	16	0.07391
TE Families 1%	16	16	0.007525
TE Families 5%	16	16	0.006287

Supplementary Materials

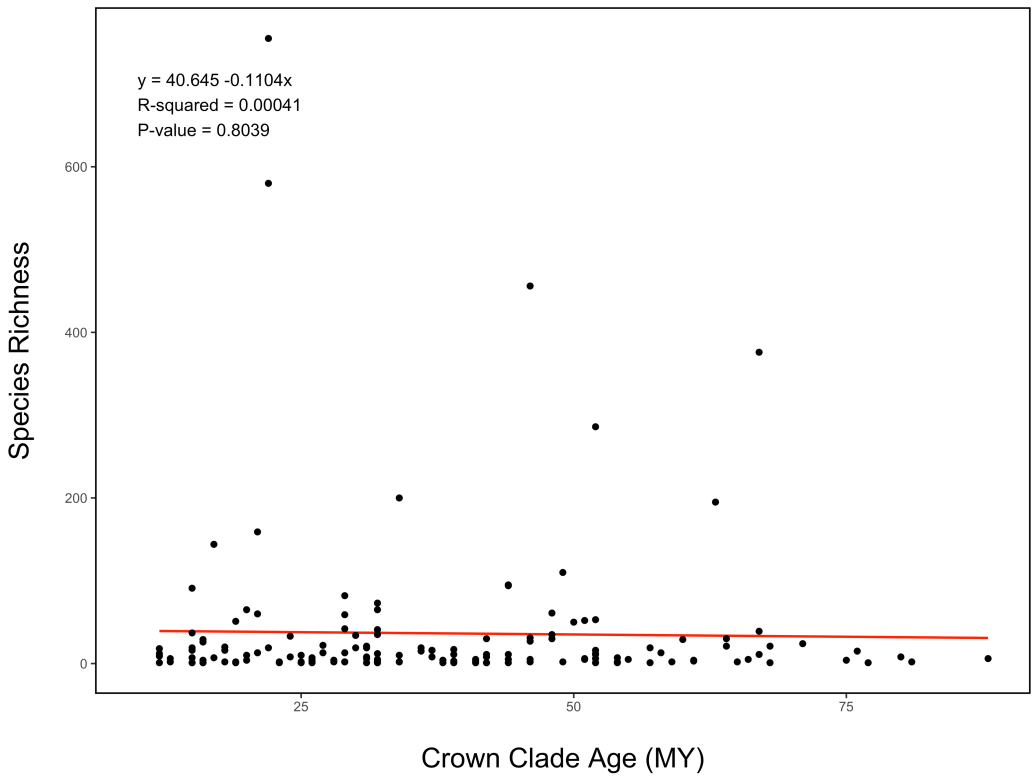


Figure S1. Relationship between number of species and crown clade age. Each point represents a family of Mammals. In red the regression line.

Table S1 Species richness and age of the orders and their species of Mammals

ORDER	FAMILY	No. SPECIES	AGE MY
Peramelemorphia	Peroryctidae	12	12
Peramelemorphia	Peramelidae	18	12
Cetartiodactyla	Balaenidae	4	12
Cetartiodactyla	Balaenopteridae	9	12
Cetartiodactyla	Eschrichtiidae	1	12
Cetacea	Phocoenidae	6	13
Cetacea	Monodontidae	2	13
Primates	Hominidae	7	14
Primates	Hylobatidae	19	14
Carnivora	Otariidae	16	15
Carnivora	Odobenidae	1	15
Cetartiodactyla	Neobalaenidae	1	15
Cetacea	Delphinidae	37	15
Rodentia	Echimyidae	91	15
Rodentia	Myocastoridae	1	15
Cetacea	Pontoporiidae	1	16
Cetacea	Iniidae	4	16
Primates	Atelidae	26	16
Primates	Cebidae	29	16
Rodentia	Hydrochoeridae	2	17
Rodentia	Cavidae	16	17
Artiodactyla	Bovidae	144	18
Artiodactyla	Moschidae	7	18
Artiodactyla	Cervidae	51	19
Artiodactyla	Antilocapridae	1	19
Artiodactyla	Giraffidae	2	19
Rodentia	Capromyidae	20	19
Diprotodontia	Macropodidae	65	21
Diprotodontia	Potoroidae	10	21
Rodentia	Octodontidae	13	21
Rodentia	Ctenomydae	60	21
Primates	Cercopithecidae	159	21
Rodentia	Cricetidae	580	22
Rodentia	Muridae	755	22
Carnivora	Phocidae	19	23
Rodentia	Thryonomidae	2	23
Rodentia	Petromuridae	1	23
Carnivora	Hespertidae	33	24
Carnivora	Eupleridae	8	24
Cetacea	Kogiidae	2	25
Cetacea	Physeteridae	1	25
Rodentia	Abrocomidae	10	25
Rodentia	Chinchilidae	7	26
Rodentia	Dinomydae	1	26

Cetacea	Ziphiidae	22	27
Rodentia	Dasyproctidae	13	27
Pilosa	Bradypodidae	4	27
Pilosa	Megalonychidae	2	27
Carnivora	Hyaenidae	4	28
Didelphimorphia	Didelphidae	82	29
Peramelemorphia	Thylacomyidae	2	29
Carnivora	Felidae	42	29
Carnivora	Prionodontidae	2	29
Carnivora	Procyonidae	13	29
Carnivora	Mustelidae	59	29
Cetacea	Platanistidae	2	29
Rodentia	Cuniculidae	2	29
lagomorpha	Ochotonidae	30	30
Primates	Cheirogaleidae	34	30
Primates	Indriidae	19	30
Didelphimorphia	Caluromyidae	5	31
Carnivora	Ailuridae	1	31
Artiodactyla	Tayassuidae	8	31
Artiodactyla	Suidae	19	31
Rodentia	Calomyscidae	6	31
Rodentia	Geomyidae	41	31
Rodentia	Heteromyidae	65	31
Primates	Lemuridae	21	31
Dasyuromorphia	Dasyuridae	73	32
Dasyuromorphia	Myrmecobiidae	1	32
Diprotodontia	Hypsiprymnodontidae	1	32
Carnivora	Viverridae	35	32
Carnivora	Mephitidae	12	32
Sirenia	Trichechidae	3	32
Sirenia	Dugongidae	2	32
Chiroptera	Noctilionidae	2	33
Chiroptera	Furipteridae	2	33
Chiroptera	Mormoopidae	10	34
Chiroptera	Phyllostomidae	200	34
Primates	Galagidae	19	36
Primates	Lorisidae	15	36
Rodentia	Erethizontidae	16	37
Monotremata	Tachyglossidae	4	38
Monotremata	Ornithorhynchidae	1	38
Carnivora	Ursidae	8	38
Rodentia	Bathyergidae	16	38
Diprotodontia	Vombatidae	3	39
Diprotodontia	Phascolarctidae	1	39
Diprotodontia	Pseudocheiridae	17	39
Diprotodontia	Petauridae	11	39

Chiroptera	Thyropeteridae	5	40
Carnivora	Nandiniidae	1	41
Pilosa	Myrmecophagidae	3	41
Pilosa	Cyclopedidae	1	41
Diprotodontia	Tarsipedidae	1	42
Artiodactyla	Tragulidae	8	42
Rodentia	Spalacidae	30	42
Rodentia	Ctenodactylidae	11	42
Rodentia	Diatomidae	1	42
Chiroptera	Hipposideridae	94	44
Chiroptera	Rhinolophidae	95	44
Chiroptera	Megadermatidae	5	44
Chiroptera	Craseonycteridae	1	44
Rodentia	Hystricidae	11	44
Diprotodontia	Acrobatidae	2	46
Diprotodontia	Phalangeridae	27	46
Diprotodontia	Burramyidae	5	46
Chiroptera	Vespertilionidae	456	46
Chiroptera	Miniopteridae	31	46
Carnivora	Canidae	35	48
Chiroptera	Mystacinidae	2	48
Chiroptera	Molossidae	110	49
lagomorpha	Leporidae	61	49
Rodentia	Dipodidae	52	51
Chiroptera	Natalidae	12	52
Chiroptera	Nycteridae	16	52
Chiroptera	Emballonuridae	53	52
Chiroptera	Rhinopomatidae	6	52
Perissodactyla	Rhinoceratidae	6	52
Perissodactyla	Tapiridae	5	52
Rodentia	Aplodontidae	1	52
Rodentia	Sciuridi	286	52
Primates	Daubentonidae	11	53
Chiroptera	Myzopodidae	2	54
Rodentia	Anomaluridae	7	54
Rodentia	Pedetida	1	54
Artiodactyla	Hippopotamidae	5	55
scandentia	Tupaiidae	19	57
scandentia	Ptilocercidae	1	57
Perissodactyla	Equidae	13	58
Rodentia	Castoridae	2	58
Rodentia	Gliridae	29	60
Proboscidae	Elephantidae	3	61
Hyracoidae	Procavidae	4	61
Chiroptera	Pteropodidae	195	62
Afrosoricidae	Tenrecidae	30	64

Afrosoricidae	Chysochloridae	21	64
Notoryctemorphia	Notoryctidae	2	65
Artiodactyla	Camelidae	5	65
Insectivora	Talpidae	39	67
Insectivora	Soricidae	376	67
Primates	Tarsidae	11	67
Cingulata	Dasypodidae	21	68
Microbiotheria	Microbiotheriidae	1	69
Insectivora	Erinaceidae	24	71
Insectivora	Solenodontidae	4	75
Macroscelidea	Macroscelididae	15	76
Tubulidentata	Orycteropodidae	1	77
Pholydotae	Manidae	8	80
	Cynocephalidae	2	81
Paucituberculata	Caenolestidae	6	87

Table S2 Results of the correlation test (Spearman method) between the rate of speciation and TE activity among the mammalian families

PARAMETER	**S**	**P-value**
1% DI	2502.2	0.0399
5% DI	2202.9	0.001168
TE Families 1%	1977.9	0.004446
TE Families 5%	2310.1	0.01958

Table S3 Results of the linear regression models between the Rate of Speciation (RS) and TE activity among the mammalian families

PARAMETER	**ANGULAR COEFFICIENT**	**R^2**	**P-value**
DI 1%	48.17	0.2617	0.004566
DI 5%	889.9	0.1662	0.02814
TE Families 1%	0.5733	0.5101	1,353 e-5
TE Families 5%	2.3843	0.4314	0.0001089