





Article

Transposable Elements in Bats Show Differential Accumulation Patterns Determined by Class and Functionality

Nicole S. Paulat ¹, Erin McGuire ², Krishnamurthy Subramanian ^{2,3}, Austin B. Osmanski ¹, Diana D. Moreno-Santillán ¹, David A. Ray ^{1,*} and Jinchuan Xing ^{2,3,*}

¹ Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA

² Department of Genetics, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

³ Human Genetics Institute of New Jersey, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

* Correspondence: david.4.ray@gmail.com (D.A.R.); jinchuan.xing@rutgers.edu (J.X.)

Abstract: Bat genomes are characterized by a diverse transposable element (TE) repertoire. In particular, the genomes of members of the family Vespertilionidae contain both active retrotransposons and active DNA transposons. Each TE type is characterized by a distinct pattern of accumulation over the past ~40 million years. Each also exhibits its own target site preferences (sometimes shared with other TEs) that impact where they are likely to insert when mobilizing. Therefore, bats provide a great resource for understanding the diversity of TE insertion patterns. To gain insight into how these diverse TEs impact genome structure, we performed comparative spatial analyses between different TE classes and genomic features, including genic regions and CpG islands. Our results showed a depletion of all TEs in the coding sequence and revealed patterns of species- and element-specific attraction in the transcript. Trends of attraction in the distance tests also suggested significant TE activity in regions adjacent to genes. In particular, the enrichment of small, non-autonomous TE insertions in introns and near coding regions supports the hypothesis that the genomic distribution of TEs is the product of a balance of the TE insertion preference in open chromatin regions and the purifying selection against TEs within genes.

Keywords: transposable elements; genome evolution; insertion preference



Citation: Paulat, N.S.; McGuire, E.; Subramanian, K.; Osmanski, A.B.; Moreno-Santillán, D.D.; Ray, D.A.; Xing, J. Transposable Elements in Bats Show Differential Accumulation Patterns Determined by Class and Functionality. *Life* **2022**, *12*, 1190. <https://doi.org/10.3390/life12081190>

Academic Editor: Kyudong Han

Received: 1 July 2022

Accepted: 2 August 2022

Published: 4 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Transposable elements (TEs), also known as mobile elements, are repetitive sequences of DNA that are able to mobilize in and across genomes through multiple mechanisms. TE transpositions not only introduce new TE sequences at the integration sites but can also introduce mutations because of DNA breaks/repairs during this process. Over time, TE transposition has the ability to change protein products, alter gene expression, and reconstruct genomic architecture [1–9].

Because of the dramatic impacts TEs can have, it is important to understand their insertion and accumulation patterns in genomes. Several hypotheses propose that TE transposition and integration can act as powerful drivers of genome evolution and facilitators of diversification [10–12]. Studies have also implicated TEs in regulatory innovations in developmental and immune pathways, as well as in human diseases [13–19]. Recent advances in sequencing technology and genomic databases allowed studies to gain detailed insights into the deep and widespread influence of TEs [20,21].

However, the spatial accumulation patterns of TEs in mammalian genomes have yet to be well established, although recent studies show TEs contribute to both 3D chromatin structure and the transcriptional regulation of nearby genes [7–9]. One hypothesis of TE accumulation states that TE activity is selected against in coding regions because of its potential destruction of genome functionality and integrity [22]. The converse school of thought is that TE activity is selected for in expressed regions because the loose chromatin

structure allows for transposition mechanisms, and the “wound” nucleosome state inhibits transposition [23].

TE diversity across species can be substantial, and different classes of TEs also have distinct transposition mechanisms. These differences may impact how the insertion regions are affected [24,25]. Class I elements (retrotransposons) replicate using a copy-and-paste mechanism: the element is transcribed into an RNA intermediate (‘copy’), which is then reverse-transcribed and inserted (‘paste’) in another location in the genome [25]. Class II elements, in particular TIR-like DNA transposons, on the other hand, ‘cut’ themselves out of the genome with a TE-encoded transposase and insert (‘paste’) into a new location. Finally, rolling circle (RC) transposons such as Helitrons (also Class II) mobilize in a replicative fashion distinct from their fellow Class II elements and from retrotransposons [25]. The downstream impacts of these three classes could therefore be distinct. For example, retrotransposition of long interspersed elements (LINEs) and short interspersed elements (SINEs) requires two sequential single-strand nicks in the target site DNA for insertion, while the transposition of many DNA transposons requires double-strand cuts to at the target site [26,27]. Like LINEs, RC transposons utilize single nicks but across multiple DNA strands. Thus, the scale of impact is almost certain to vary across elements and across species, but the extent of these differences, and their related mechanisms, are not yet clear. This is especially true for mammals as the vast majority of TE research in this clade has been conducted in species that do not harbor active Class II elements.

Another area of interest is the non-random distribution of TEs across a genome. Several factors impact where TEs accumulate. Different TEs have different target site preferences, and even those with the same preference can accumulate in different genomic regions. For example, in mammals, retrotransposons such as LINEs and SINEs prefer AT-rich target sites, but LINEs accumulate in AT-rich regions while SINEs tend to accumulate in GC-rich regions [28,29]. There are many possible reasons for this pattern, including TE size, how effectively host defenses interfere with transposition, whether the target site is in open or closed chromatin, or selection against non-homologous recombination to remove existing insertions [23,28,30]. However, studies examining the factors affecting TE distribution patterns mainly focused on a few plant and animal models (*Arabidopsis*, maize, *Drosophila*, mouse, and human), so the universality of such patterns remains to be evaluated.

In this study, we examined the correlation between annotated TEs and functional regions of the genome to test the hypothesis that TE insertions are selected against in genic regions. We also analyzed the activity of TE insertions relative to their age. We leveraged the increased TE diversity observed in bats to determine if Class II elements exhibit distinct patterns relative to the more extensively studied Class I elements. The data show a depletion of all TEs in the coding sequence and reveal patterns of species- and element-specific attraction in the transcript region. Trends of attraction in the absolute and relative distance tests suggest significant TE activity in regions adjacent to genes. These data suggest that the accumulation of TEs is influenced by selective pressure against TEs in protein-coding sequences and against insertions in nucleosomal DNA. The type of TE also contributes to the selection pressure. The enrichment of small, non-autonomous TEs in introns and near coding regions supports the hypothesis that genomic locations of TEs are the product of a balance of the two competing selective pressures.

2. Materials and Methods

2.1. Species Selection and TE Annotation

Genomes of seven species of bats were included in this study (Table S1). Genome assemblies and gene annotation files of six bat species were downloaded from the Bat1K project (<https://bds.mpi-cbg.de/hillerlab/Bat1KPilotProject/>, accessed on 1 April 2022). The *Myotis lucifugus* genome assembly (GCA_000147115.1) and gene annotation were obtained from the UCSC genome browser (<https://hgdownload.soe.ucsc.edu/goldenPath/myoLuc2/bigZips/myoLuc2.fa.gz>, accessed on 1 April 2022; <https://hgdownload.soe.ucsc.edu/goldenPath/myoLuc2/bigZips/genes/myoLuc2.ncbiRefSeq.gtf.gz>, accessed on

1 April 2022). The CpG islands in each genome were identified based on the genome assembly using a hidden Markov model implemented in the program makeCGI (v 1.3.4) [31].

TE insertions were annotated with RepeatMasker v4.0.9 using a custom TE library [32]. This TE library is a combination of mammalian TE sequences that were available from RepBase as well as those manually curated by the Zoonomia project. The custom TE library is available as a supplemental file at <https://doi.org/10.7282/00000212> (accessed on 1 August 2022), and all consensus have been submitted to DFAM [33].

Raw RepeatMasker output was reformatted to BED format and split by TE class. TE classes were defined as LINE, SINE, LTR (long terminal repeat), DNA, and RC. TEs shorter than 100 base pairs (bps) were excluded from further analysis. A previous study showed that RC activities were exclusive to vesper bats among bat species in the past ~40 million years [34]. The small number of RCs annotated by RepeatMasker in species outside of the Vespertilionidae linkage were likely to be false positives and were therefore excluded from the analysis. Old and young TE subsets were determined based on approximate age. For the young group, TEs younger than ~25 million years (Mys) were selected; for the old group, TEs older than 40 Mys were selected. The age of the TEs was calculated based on species-specific mutation rates (Table S1) determined by consensus phylogeny branch lengths from Foley et al. [35] and median divergence estimates from TimeTree [36]. The TE selection was performed using filter_beds.py (https://github.com/davidaray/bioinfo_tools/blob/master/filter_beds.py, accessed on 1 April 2022).

2.2. TE Spatial Distribution Analysis

The GenometriCorr package's suite (<http://genometricorr.sourceforge.net/>, accessed on 1 April 2022) was used to evaluate whether the spatial distribution of each TE group is independent of the positions of genes [37]. The three statistical tests were the Jaccard measure test, absolute distance test, and relative distance test. The significance of the tests was obtained through 100 permutations. The overall results of each test, based on all chromosomes with both TE and gene annotations, were used to determine if each TE group had a non-random spatial distribution in the bat genomes.

3. Results

3.1. TE Composition in Seven Bat Species

For this study, we selected seven bat species (Figure 1). Two features of bat lineages among mammals allowed us to compare the distribution of class I and class II TEs in a unique fashion. First, unlike most mammals and most other bats, DNA and RC transposons are active within the vesper bat lineage (family Vespertilionidae) [38–41]. Second, the pteropodid lineage of megabats (family Pteropodidae) has experienced a nearly complete cessation of activity from all types of TEs over the past 40 Mys [34,42,43].

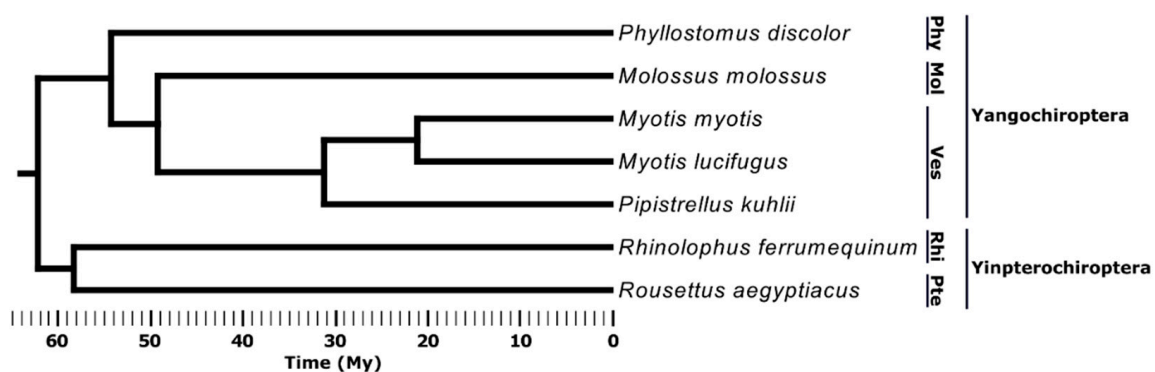


Figure 1. Phylogenetic tree of species in this study (based on [32]). Bat suborder is shown at far right, and family is indicated by three-letter abbreviations: Phy = Phyllostomidae, Mol = Molossidae, Ves = Vespertilionidae, Rhi = Rhinolophidae, Pte = Pteropodidae.

First, we determined the TE composition in the seven bat genomes (Figure 2, Table S1). LINES and SINES are the most abundant TE classes in the seven species, a typical characteristic of mammals [34]. The numbers of SINES and LINES are the highest in the *Mo. molossus* genome (1,158,863 SINES, 1,092,308 LINES; Table S1). The composition of the genomes is relatively uniform for LINE, LTR, and DNA transposons but are variable for SINE and RC elements. For example, the genomes *My. myotis*, *My. lucifugus*, and *Pi. kuhlii* have about half a million RC insertions, while there are almost no RC insertions in other species.

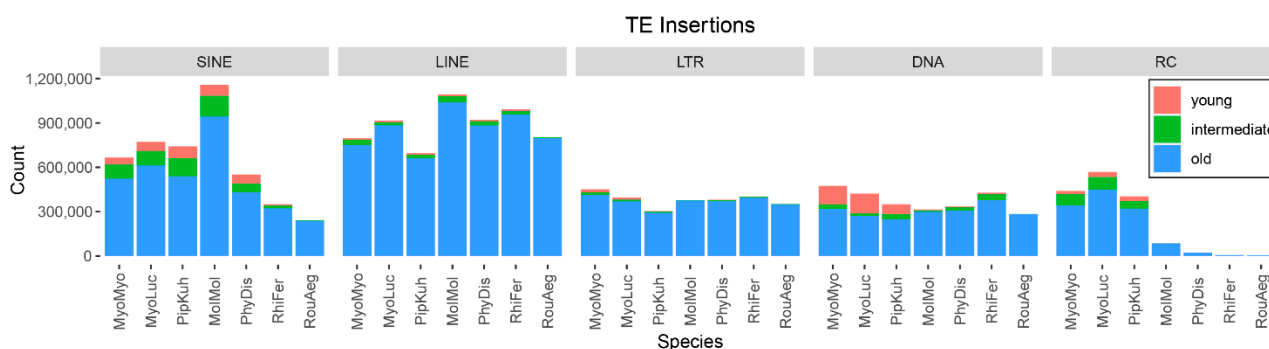


Figure 2. The number of annotated TEs for each species. Young: elements < 25 Mys old; Old: elements > 40 Mys old; Intermediate: elements between 25 and 40 Mys old. Species abbreviations: MyoMyo: *My. myotis*; MyoLuc: *My. lucifugus*; PipKuh: *Pi. kuhlii*; MolMol: *Mo. molossus*, PhyDis: *Ph. discolor*; RihFer: *Rh. ferrumequinum*; RouAeg: *Ro. aegyptiacus*.

Next, we examined the TE insertion dynamics in the seven genomes by counting the number of young (<25 Mys), intermediate (between 25 and 40 Mys), and old (>40 Mys) TEs in each genome (Figure 2). The proportion of young insertions is small, ranging from almost no activity in *Ro. aegyptiacus* to about 8% of all TEs in vesper bats. DNA transposons are the most active young elements in *My. myotis* and *My. lucifugus*, making up 55% and 52% of young insertions in their genomes, respectively. In *Ph. discolor*, *Mo. molossus*, and *Pi. kuhlii*, SINES are the most active young elements and account for 80%, 80%, and 41% of all young insertions, respectively. The *Rh. ferrumequinum* genome contains a similar number of young SINE, LINE, and DNA elements (~10,000). As described previously, *Ro. aegyptiacus* have almost no recent TE activity.

We then compared the size distribution of different TEs (Figure 3). Most SINES, RCs, and DNA transposons are less than 300 bps in length. While the full length of LINE and LTR elements are 6–8 kbps in size, LINES and LTRs in bat genomes showed a wide distribution. This is expected given that most LTR elements are solo-LTRs [44], while most LINES are truncated [45].

3.2. TE Distribution Relative to Genic Regions and CpG Islands

We determined the relationship between genic regions, CpG islands (CGIs), and TE distributions across the genomes. For genic regions, we included the transcript, coding sequence (CDS), and the start codon position in our analysis. The transcript region is the broadest category tested, including introns, CDS, and untranslated regions. Compared to transcripts, the CDS and start codon represent the more conserved protein-coding regions. CGIs are regions of the genome with high levels of CpG dinucleotides and are generally associated with promoters [46]. We will collectively refer to the transcript, CDS, start codon, and CGI as genic features in the following text. For TEs, we tested three classes of retrotransposons (SINE, LINE, and LTR) and two classes of DNA transposons (DNA and RC). We focused on young (<25 Mys) and old (>40 Mys) TEs to determine the effect of insertion site preference and the effect of purifying selection against TEs on the TE genomic distribution, respectively. We applied Jaccard, absolute distance, and relative distance tests to examine the relationships between TEs and genic features. The Jaccard test determines

the overlap between two sets of features (e.g., transcript and SINE) based on the ratio of the intersection to the union of the features. The absolute distance test measures the absolute minimum distance between two sets of features, while the relative distance test measures the average distance between the midpoints of two sets of features.

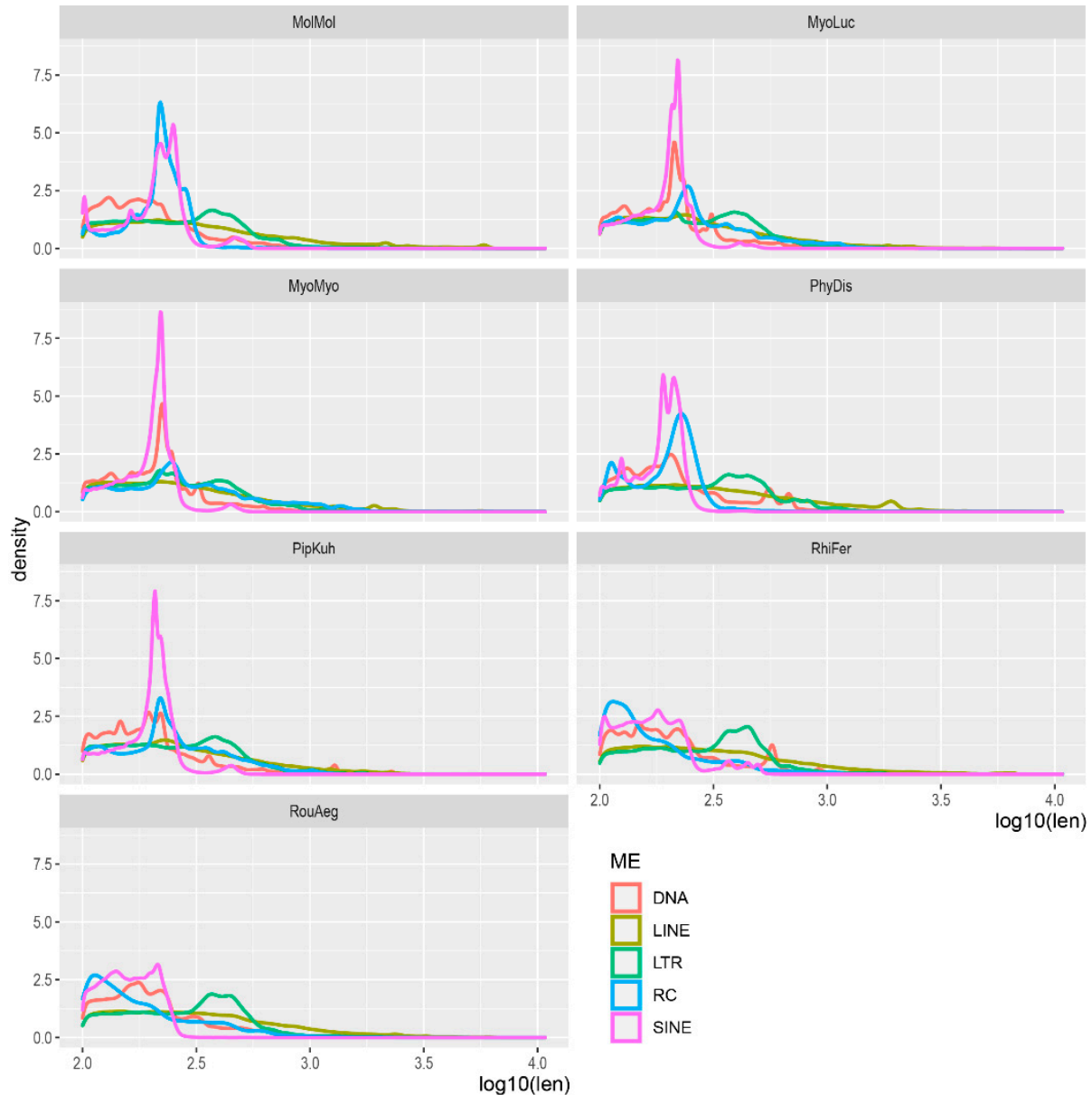


Figure 3. Size distribution of TEs in each species. The x-axis represents TE size (bps) on a log₁₀ scale. Species abbreviations: MyoMyo: *My. myotis*; MyoLuc: *My. lucifugus*; PipKuh: *Pi. kuhlii*; MolMol: *Mo. molossus*, PhyDis: *Ph. discolor*; RhiFer: *Rh. ferrumequinum*; RouAeg: *Ro. aegyptiacus*.

Figure 4 summarizes the correlations between TEs and genic features. Among old TEs (>40 Mys), the Jaccard test showed an overall pattern of repulsion among all TEs and gene features (Figure 4A). The only exceptions are SINE and DNA elements, which show enrichment of insertions (i.e., attraction) in the transcript of some species. The vast majority of SINE and DNA elements are non-autonomous and contain no open reading frames (ORFs). For the two types of autonomous retrotransposons, LTR and LINE, the Jaccard test showed depletion within all genic features in all species. Similarly, in vesper bat species that contain a high number of RC insertions, the Jaccard test showed depletion of RCs within all genic features. In contrast, the relative distance test showed an overall

attraction between old TEs and gene features. The results of the absolute distance tests were intermediate between the Jaccard and relative distance test: while LINE, LTR, and RC elements showed depletion in genic features, SINE and DNA elements showed attraction patterns across the species, except for DNA elements and CGI in *Rh. ferrumequinum*. Taken together, these results suggest that, in general, old TEs are depleted within genic features but are in proximity of these features.

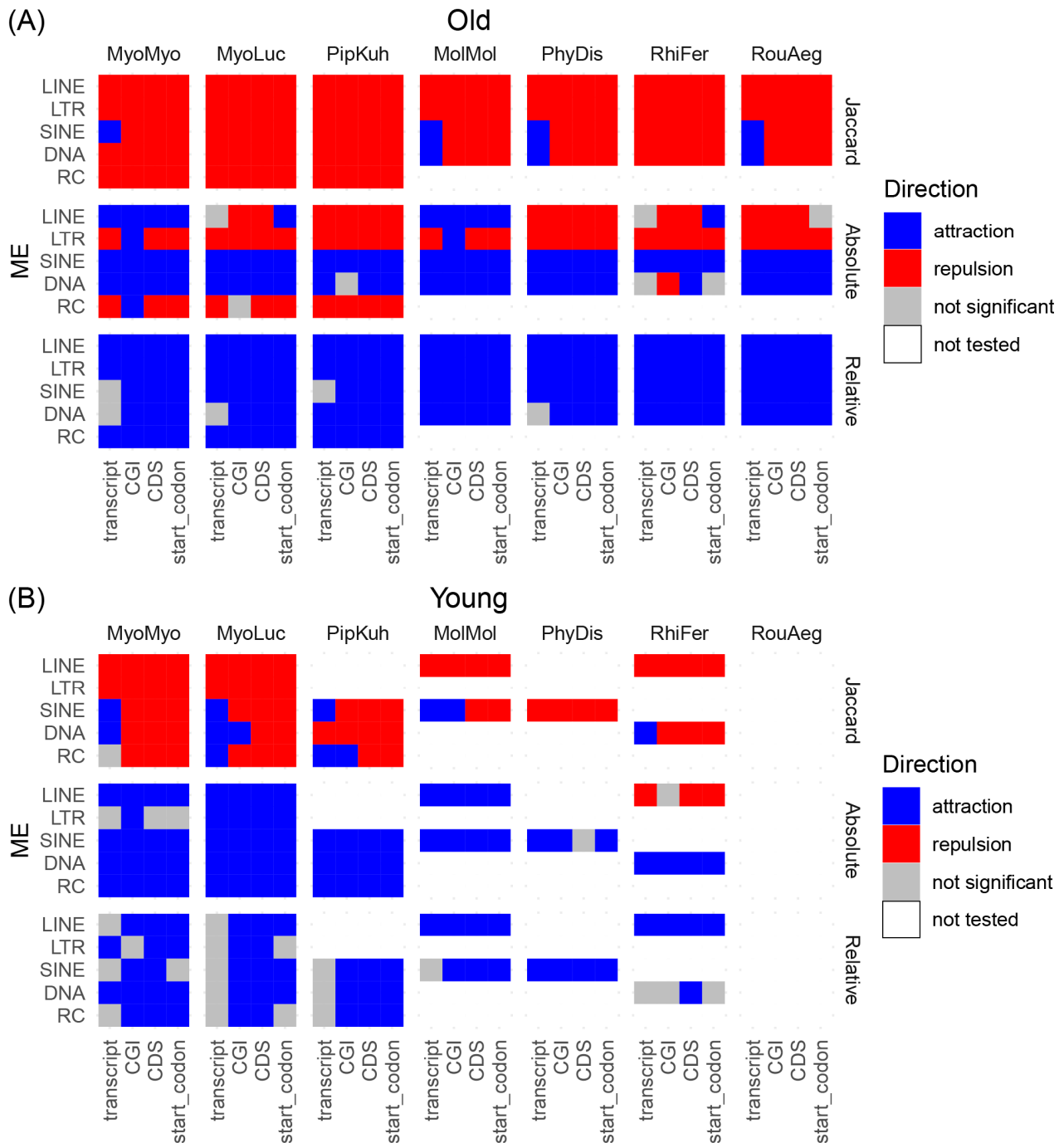


Figure 4. Relationship between TEs and genic features. Results from the Jaccard, absolute distance, and relative distance tests for old (A) and young (B) TEs. Minimum cutoff of 100 bp length for TE insertion, and 10,000 insertions per element per species. Attraction is indicated in blue, repulsion in red, no-significant in grey, and not-tested in white. Species abbreviations: MyoMyo: *My. myotis*; MyoLuc: *My. lucifugus*; PipKuh: *Pi. kuhlii*; MolMol: *Mo. molossus*, PhyDis: *Ph. discolor*; RhiFer: *Rh. ferrumequinum*; RouAeg: *Ro. aegyptiacus*.

For young TEs (<25 Mys), because of the general low recent TE activities in bat species besides the vesper bats (Figure 2), we did not have enough insertions to test the interaction between some TEs and genic features. For groups that are large enough to be tested (i.e., >10,000 elements), the trends are generally consistent with old TEs in the Jaccard test: SINE and DNA elements are enriched within transcripts, while LINE and LTR elements are depleted in all genic features (Figure 4B). Interestingly, young RC elements in *My. lucifugus* and *Pi. kuhlii* showed attraction to the transcript, and DNA elements in *My. lucifugus* and SINEs in *Mo. molossus* showed attraction to CGI in the Jaccard test. For the absolute and relative distance tests, most of the tests showed enrichment of TEs within genic features, except for LINE elements in *Rh. ferrumequinum* (Figure 4B, Table S2).

4. Discussion

In this study, we examined the spatial relationships between all TE types and genic features in a variety of bats. Bats were chosen due to several clades that show unique TE characteristics when compared to other mammals. In particular, the recent activity of DNA transposons (Class II TEs) in vesper bats provides an opportunity to broaden the observations when compared to previous studies that had focused on mammals that have primarily experienced retrotransposon (Class I TEs) activities.

The near-unanimous attraction in the relative distance test demonstrates that TE insertions are closer to genic features than a random distribution. This result suggests a disproportionate accumulation of TEs in areas surrounding or within genes and it is consistent with the hypothesis proposed by Bennetzen [23], which states the openness of chromatin around genes, and especially around promoters, may play a role in promoting TE insertion in those regions.

In contrast, in the Jaccard test, where the intersection of features was evaluated, SINE and DNA transposons show differing results against transcripts compared to LINE and LTR elements. Both young and old LINE and LTR elements exhibit strong depletion in all genic features, suggesting selection against these insertions near genes, even soon after the insertion event. Conversely, both young and old SINEs and DNA transposons showed attraction against transcripts in several species. This is broadly consistent with trends shown in human and mouse genomes [28,30]. This different pattern may be explained by several differences among these TE groups. LINEs and LTRs are autonomous retrotransposons, and the full-length elements are large in size (6–8 kbps). As autonomous elements, LINEs and LTRs also harbor ORFs and numerous other functional features (e.g., promoters, polyadenylation sites, etc.). Therefore, the presence of a LINE or LTR insertion within or near a gene could have a dramatic impact on gene expression. On the other hand, SINE, DNA, and RC insertions are mostly small (~300 bps) and non-autonomous, with no ORFs and fewer potentially functional units than LINE/LTR insertions. The combination of smaller size and fewer potentially disruptive functional sequences may result in a weaker selective pressure against these types of insertions compared to LINE and LTR elements. To determine the impact of size on the distribution of LINE and LTR elements, we separated LINEs and LTRs into large (>500 bps) and small (≤ 500 bps) groups and tested their distributions within and outside of transcripts. The test showed that there are significantly fewer large LINEs/LTRs than small LINEs/LTRs within transcripts (Table S3, chi-square test with Yates' continuity correction $p < 10^{-8}$). These observations support the hypothesis of selection against large elements transposing into genic regions and the difficulty of successfully removing small elements very close to genes without impacting gene or regulatory region functions [22,28,47] (alternative explanation in [48]). Unlike human and mouse genomes, we also showed that DNA and young RC transposons have a similar distribution pattern to SINEs. These results may be due to the importance of chromatin accessibility for the insertion of these elements. Alternatively, shared distribution patterns of SINE and DNA transposons might suggest that functional impact, rather than the integration mechanism, plays a bigger role in determining the spatial distribution of TEs in the genome.

This body of data provides evidence to support the hypothesis that TE distribution is selected against in coding regions and enriched for upstream and downstream of genes. Further research is needed, given that our analyses are limited to bat species. In addition, this work did not extend to examining more complex relationships between genome structure, such as Z-DNA repeat sequences and TE integrations. Future studies can employ the same methods across the genomes of other bats and mammalian species to see if more nuanced patterns emerge between TE activity and genic regions. Continued research in the attraction and repulsion of TEs and genes relative to the age of the insertion may elucidate evolutionary relationships and TE influence on genome architecture.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/life12081190/s1>, Table S1: Genomic features; Table S2: GenometriCorr results; Table S3: Distribution of large and small LINES/LTRs in transcripts.

Author Contributions: Conceptualization, D.A.R. and J.X.; methodology, N.S.P., E.M. and K.S.; formal analysis, N.S.P., E.M., K.S., A.B.O., D.D.M.-S., D.A.R. and J.X.; writing—original draft preparation, N.S.P., E.M., D.A.R. and J.X.; supervision, D.A.R. and J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Foundation, grant numbers 2032006, 1838283, and 1355176, and by the Texas Tech University College of Arts and Sciences.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome annotation files for seven bat species used for the study, including gene annotation, CGI annotation, and TE annotation, are available at Rutgers University Community Repository (DOI: <https://doi.org/10.7282/00000212>, accessed on 1 August 2022).

Acknowledgments: The TTU High-Performance Computing Cluster (<https://www.depts.ttu.edu/hpcc/>, accessed on 1 April 2022) provided computational resources and technical support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McClintock, B. Chromosome organization and genic expression. *Cold Spring Harb. Symp. Quant. Biol.* **1951**, *16*, 13–47. [[CrossRef](#)] [[PubMed](#)]
2. Plasterk, R.H. The origin of footprints of the Tc1 transposon of *Caenorhabditis elegans*. *EMBO J.* **1991**, *10*, 1919–1925. [[CrossRef](#)] [[PubMed](#)]
3. Jiang, N.; Bao, Z.; Zhang, X.; Eddy, R.S.; Wessler, R.S. Pack-MULE transposable elements mediate gene evolution in plants. *Nature* **2004**, *431*, 569–573. [[CrossRef](#)] [[PubMed](#)]
4. Sen, S.K.; Han, K.; Wang, J.; Lee, J.; Wang, H.; Callinan, A.P.; Dyer, M.; Cordaux, R.; Liang, P.; Batzer, A.M. Human genomic deletions mediated by recombination between Alu elements. *Am. J. Hum. Genet.* **2006**, *79*, 41–53. [[CrossRef](#)] [[PubMed](#)]
5. Xing, J.; Zhang, Y.; Han, K.; Salem, A.H.; Sen, K.S.; Huff, D.C.; Zhou, Q.; Kirkness, F.E.; Levy, S.; Batzer, A.M.; et al. Mobile elements create structural variation: Analysis of a complete human genome. *Genome Res.* **2009**, *19*, 1516–1526. [[CrossRef](#)]
6. Wells, J.N.; Feschotte, C. A Field Guide to Eukaryotic Transposable Elements. *Annu Rev. Genet.* **2020**, *54*, 539–561. [[CrossRef](#)] [[PubMed](#)]
7. Cao, X.; Zhang, Y.; Payer, L.M.; Lords, H.; Steranka, J.P.; Burns, K.H.; Xing, J. Polymorphic mobile element insertions contribute to gene expression and alternative splicing in human tissues. *Genome Biol.* **2020**, *21*, 185. [[CrossRef](#)]
8. Zhang, Y.; Li, T.; Preissl, S.; Amaral, M.L.; Grinstein, J.D.; Farah, E.N.; Destici, E.; Qiu, Y.; Hu, R.; Lee, A.Y.; et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **2019**, *51*, 1380–1388. [[CrossRef](#)]
9. Diehl, A.G.; Ouyang, N.; Boyle, A.P. Transposable elements contribute to cell and species-specific chromatin looping and gene regulation in mammalian genomes. *Nat. Commun.* **2020**, *11*, 1796. [[CrossRef](#)]
10. Oliver, K.R.; Greene, W.K. Mobile DNA and the TE-Thrust hypothesis: Supporting evidence from the primates. *Mob. DNA* **2011**, *2*, 8. [[CrossRef](#)]
11. Oliver, K.R.; Greene, W.K. Transposable elements and viruses as factors in adaptation and evolution: An expansion and strengthening of the TE-Thrust hypothesis. *Ecol. Evol.* **2012**, *2*, 2912–2933. [[CrossRef](#)] [[PubMed](#)]
12. Jurka, J.; Bao, W.; Kojima, K.K. Families of transposable elements, population structure and the origin of species. *Biol. Direct* **2011**, *6*, 44. [[CrossRef](#)] [[PubMed](#)]

13. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **2017**, *18*, 71–86. [[CrossRef](#)]
14. Gerdes, P.; Richardson, S.R.; Mager, D.L.; Faulkner, G.J. Transposable elements in the mammalian embryo: Pioneers surviving through stealth and service. *Genome Biol.* **2016**, *17*, 100. [[CrossRef](#)]
15. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **2016**, *351*, 1083–1087. [[CrossRef](#)]
16. Lynch, V.J.; Nnamani, M.C.; Kapusta, A.; Brayer, K.; Plaza, S.L.; Mazur, E.C.; Emera, D.; Sheikh, S.Z.; Grutzner, F.; Bauersachs, S.; et al. Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep.* **2015**, *10*, 551–561. [[CrossRef](#)]
17. Friedli, M.; Trono, D. The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* **2015**, *31*, 429–451. [[CrossRef](#)]
18. Hancks, D.C.; Kazazian, H.H., Jr. Active human retrotransposons: Variation and disease. *Curr. Opin. Genet. Dev.* **2012**, *22*, 191–203. [[CrossRef](#)]
19. Xing, J.; Wang, H.; Belancio, V.P.; Cordaux, R.; Deininger, P.L.; Batzer, M.A. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 17608–17613. [[CrossRef](#)]
20. Xing, J.; Witherspoon, D.J.; Jorde, L.B. Mobile element biology: New possibilities with high-throughput sequencing. *Trends Genet.* **2013**, *29*, 280–289. [[CrossRef](#)]
21. Zoonomia, C. A comparative genomics multitool for scientific discovery and conservation. *Nature* **2020**, *587*, 240–245. [[CrossRef](#)] [[PubMed](#)]
22. Medstrand, P.; van de Lagemaat, L.N.; Mager, D.L. Retroelement distributions in the human genome: Variations associated with age and proximity to genes. *Genome Res.* **2002**, *12*, 1483–1495. [[CrossRef](#)] [[PubMed](#)]
23. Bennetzen, J.L. Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* **2000**, *42*, 251–269. [[CrossRef](#)]
24. Sotero-Caio, C.G.; Platt, R.N.; Suh, A.; Ray, D.A. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol. Evol.* **2017**, *9*, 161–177. [[CrossRef](#)] [[PubMed](#)]
25. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)]
26. Cost, G.J.; Feng, Q.; Jacquier, A.; Boeke, J.D. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* **2002**, *21*, 5899–5910. [[CrossRef](#)] [[PubMed](#)]
27. Hickman, A.B.; Perez, Z.N.; Zhou, L.; Musingarimi, P.; Ghirlando, R.; Hinshaw, J.E.; Craig, N.L.; Dyda, F. Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol.* **2005**, *12*, 715–721. [[CrossRef](#)] [[PubMed](#)]
28. Jurka, J.; Kohany, O.; Pavlicek, A.; Kapitonov, V.V.; Jurka, M.V. Duplication, coclustering, and selection of human Alu retrotransposons. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1268–1272. [[CrossRef](#)]
29. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 1872–1877. [[CrossRef](#)]
30. Jurka, J.; Kohany, O.; Pavlicek, A.; Kapitonov, V.; Jurka, M. Clustering, duplication and chromosomal distribution of mouse SINE retrotransposons. *Cytogenet. Genome Res.* **2005**, *110*, 117–123. [[CrossRef](#)]
31. Wu, H.; Caffo, B.; Jaffee, H.A.; Irizarry, R.A.; Feinberg, A.P. Redefining CpG islands using hidden Markov models. *Biostatistics* **2010**, *11*, 499–514. [[CrossRef](#)] [[PubMed](#)]
32. Jebb, D.; Huang, Z.; Pippel, M.; Hughes, G.M.; Lavrichenko, K.; Devanna, P.; Winkler, S.; Jermini, L.S.; Skirmuntt, E.C.; Katzourakis, A.; et al. Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **2020**, *583*, 578–584. [[CrossRef](#)]
33. Hubley, R.; Finn, R.D.; Clements, J.; Eddy, S.R.; Jones, T.A.; Bao, W.; Smit, A.F.; Wheeler, T.J. The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **2016**, *44*, D81–D89. [[CrossRef](#)]
34. Osmanski, A.B.; Paulat, N.S.; Korstian, J.; Grimshaw, J.R.; Halsey, M.; Sullivan, K.A.M.; Moreno-Santillán, D.D.; Crookshanks, C.; Roberts, J.; Garcia, C.; et al. Insights into mammalian TE diversity via the curation of 200+ mammalian genome assemblies. *Science* **2022**. *submitted*.
35. Foley, N.M.; Mason, V.C.; Harris, A.J.; Bredemeyer, K.R.; Damas, J.; Lewin, H.A.; Eizirik, E.; Gatesy, J.; Zoonomia, C.; Springer, M.S.; et al. A genomic timescale for placental mammal evolution. *Science* **2022**, *in press*.
36. Kumar, S.; Stecher, G.; Suleski, M.; Hedges, S.B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **2017**, *34*, 1812–1819. [[CrossRef](#)] [[PubMed](#)]
37. Favorov, A.V.; Mularoni, L.; Cope, L.M.; Medvedeva, Y.; Mironov, A.A.; Makeev, V.; Wheelan, S.J. Exploring Massive, Genome Scale Datasets with the GenometriCorr Package. *PLoS Comput. Biol.* **2012**, *8*, e1002529. [[CrossRef](#)] [[PubMed](#)]
38. Platt, R.N., 2nd; Mangum, S.F.; Ray, D.A. Pinpointing the vesper bat transposon revolution using the *Miniopterus natalensis* genome. *Mob. DNA* **2016**, *7*, 12. [[CrossRef](#)]
39. Ray, D.A.; Feschotte, C.; Pagan, H.J.; Smith, J.D.; Pritham, E.J.; Arensburger, P.; Atkinson, P.W.; Craig, N.L. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* **2008**, *18*, 717–728. [[CrossRef](#)]
40. Ray, D.A.; Pagan, H.J.; Thompson, M.L.; Stevens, R.D. Bats with hATs: Evidence for recent DNA transposon activity in genus *Myotis*. *Mol. Biol. Evol.* **2007**, *24*, 632–639. [[CrossRef](#)]

41. Pritham, E.J.; Feschotte, C. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1895–1900. [[CrossRef](#)]
42. Yang, L.; Scott, L.; Wichman, H.A. Tracing the history of LINE and SINE extinction in sigmodontine rodents. *Mob. DNA* **2019**, *10*, 22. [[CrossRef](#)]
43. Wichman, H.A.; Scott, L.; Howell, E.K.; Martinez, A.R.; Yang, L.; Baker, R.J. Flying Around in the Genome: Characterization of LINE-1 in Chiroptera. *Spec. Publ. Tex. Tech Univ. Mus.* **2019**, *71*, 379–392.
44. Hughes, J.F.; Coffin, J.M. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: Implications for human and viral evolution. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1668–1672. [[CrossRef](#)]
45. Szak, S.T.; Pickeral, O.K.; Makalowski, W.; Boguski, M.S.; Landsman, D.; Boeke, J.D. Molecular archeology of L1 insertions in the human genome. *Genome Biol.* **2002**, *3*, research0052.1–0052.18. [[CrossRef](#)]
46. Saxonov, S.; Berg, P.; Brutlag, D.L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1412–1417. [[CrossRef](#)] [[PubMed](#)]
47. Abrusán, G.; Krambeck, H.-J. The Distribution of L1 and Alu Retroelements in Relation to GC Content on Human Sex Chromosomes Is Consistent with the Ectopic Recombination Model. *J. Mol. Evol.* **2006**, *63*, 484. [[CrossRef](#)]
48. Kvikstad, E.M.; Makova, K.D. The (r)evolution of SINE versus LINE distributions in primate genomes: Sex chromosomes are important. *Genome Res.* **2010**, *20*, 600–613. [[CrossRef](#)]