The PIWI/piRNA response is relaxed in a rodent that lacks mobilizing transposable elements

Michael W. Vandewege[1], Roy N. Patt II[2], Dana K. Merriman[3], David A. Ray[4], Federico G. Hoffmann[5,6]

[1]Department of Biology, Eastern New Mexico University
[2]Disease intervention and Prevention Program, Texas Biomedical Research Institute
[3]Department of Biology, University of Wisconsin Oshkosh
[4]Department of Biological Science, Texas Tech University
[5]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University
[6]Institute of Genomics, Biocomputing and Biotechnology, Mississippi State University

Corresponding Author:

Federico G. Hoffmann
Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology
Box 9655
Mississippi State University
Mississippi State, MS 39762
E-mail: federico.g.hoffmann@gmail.com
Office phone: 662-325-2763
Fax: 662-325-8955

Abstract

Transposable elements (TEs) have the capability to propagate throughout genomes. Mammalian genomes are typically dominated by LINE retrotransposons and their associated SINEs, and their mobilization in the germline is a challenge to genome integrity. There are genomic defenses against TE proliferation and the PIWI/piRNAs defense is among the most well understood. However, the PIWI/piRNA system has been investigated largely in animals with abundant and actively mobilizing TEs and it is unclear how the PIWI/piRNA system functions in the absence of mobilizing TEs. The thirteen-lined ground squirrel provides an excellent opportunity to examine PIWI/piRNA and TE dynamics within the context of minimal, and possibly nonexistent, TE accumulation. We sequenced RNA and small RNAs pools from the testis of juvenile and adult squirrels and compared results to TE and PIWI/piRNA dynamics in the European rabbit (*Oryctolagus cuniculus*) and house mouse (*Mus musculus*). Interestingly in squirrels, despite a lack of young insertions, TEs were still actively transcribed in comparable levels to mouse and rabbit. All three PIWI proteins were either not expressed, or only minimally expressed, prior to P8 in squirrel testis, moreover we also discovered that PIWIL4 is expressed all the way into adulthood in the squirrel. This is a crucial difference as PIWIL4 is understood to facilitate TE methylation. We present evidence the PIWI/piRNA system reduced TE expression in rabbit and mouse, but the squirrel PIWIs largely did not affect TE expression. These observations indicate that L1s do not represent a major threat to genome integrity in the squirrel genome, and therefore repression mechanisms have relaxed.

Introduction

Transposable elements (TEs) are genomic parasites that can propagate by inserting copies of themselves into the genome of their host. They can account for up to 70% of mammalian genome content (de Koning et al. 2011). Because of their ability to mobilize, TEs are powerful mutagens as novel TE insertions can disrupt exons, regulatory elements, and splice junctions, and also facilitate non-homologous recombination. As a result, TE insertions have been linked to genomic deletions, duplications, inversions, translocations, and chromosome breaks in a variety of genomes (Cheng et al. 2005; Franke et al. 2017; Platt et al. 2018). Although some TE insertions have proven to be adaptive, TEs are generally considered a serious challenge to genome integrity.

Eukaryotic genomes have therefore evolved mechanisms to restrict TE mobilization, especially in the germ line. PIWI proteins and PIWI-interacting RNAs (piRNAs) have emerged as key components in protecting the genome against the proliferation of TEs, and probably evolved in response to the challenge presented by them (Aravin, Hannon, et al. 2007; Brennecke et al. 2007; Aravin et al. 2008; Malone and Hannon 2009; Siomi et al. 2011). The general principle is that piRNAs and PIWI proteins assemble into RNA-induced silencing complexes, which go on to neutralize TE-like targets by transcript cleavage or chromatin methylation (Aravin, Sachidanandam, et al. 2007; Carmell et al. 2007; Houwing et al. 2007; Molaro et al. 2014).

Of the characterized PIWI/piRNA models, vertebrates, fruit fly, and *C. elegans* each possess independently derived sets of PIWI paralogs which associate with different sets of piRNAs (Lewis et al.; Reddien et al. 2005; Kerner et al. 2011). In mammals, mouse (*Mus musculus*) is the most studied system (reviewed in Ernst et al. (2017)). Mouse piRNAs have been categorized into two major sets, pre-pachytene and pachytene, according to when their expression begins (Aravin et al. 2006; Girard et al. 2006). Pre-pachytene piRNAs are first expressed in the early stages of spermatogenesis, are enriched for TE-like sequences, and preferentially associate with the PIWI paralogs, PIWIL2 and PIWIL4. Pre-pachytene piRNAs are thought to be largely responsible for silencing TEs. By contrast, pachytene piRNAs begin their expression when spermatocytes reach the pachytene stage, are largely derived from intergenic regions, and preferentially associate with PIWIL1. Even though pachytene piRNAs are the most abundant piRNA in adult mouse testes, their biological role is still poorly understood and current evidence suggests they play a role similar to microRNAs, silencing messenger RNAs in mouse spermiogenesis (Li et al. 2013; Gou et al. 2014).

The mouse genome encodes for three PIWI paralogs that are also differentially expressed during development and associate preferentially with piRNAs of different sizes. PIWIL2, also known as MILI, preferentially binds piRNAs 26-27 nucleotides long, begins expression at embryonic day 12.5 (E12.5) in developing testes, and is linked to the post-transcriptional silencing of TEs. PIWIL4, or MIWI2, preferentially binds to piRNAs ~28 nucleotides long, is expressed between E15.5 and post-natal day 3 (P3), and is linked to the *de novo* establishment of methylation marks in primordial germ cells (Aravin et al. 2008). Finally, PIWIL1, or MIWI,

begins expression at P14, when the pachytene stage of meiosis begins and continues through adulthood. This paralog preferentially binds to piRNAs ~30 nucleotides long and is mostly linked to spermiogenesis (Lau et al. 2006). Both PIWIL2 and PIWIL4 participate in a feed-forward loop known as the ping-pong cycle that servres to reduce the abundance of TE transcripts through piRNA-guided cleavage of TE-derived mRNAs (Aravin, Sachidanandam, et al. 2007; Aravin et al. 2008; Kuramochi-Miyagawa et al. 2008; De Fazio et al. 2011). A complete understanding of the functional roles of piRNAs is still lacking, but we know that defects in PIWI paralogs lead elevated TE activity and problems in the male germline (Carmell et al. 2007; Houwing et al. 2007; O'Donnell and Boeke 2007).

We know that TE activity and accumulation rates vary among different mammals, both in terms of the type of TEs that are active and in the level of challenge presented by them. However, most efforts at understanding piRNA/PIWI homolog biology in vertebrates have focused on the mouse model, and it is not clear whether the developmental changes in PIWI and piRNA expression observed are evolutionarily conserved or how piRNA repertoires react to changes in the TE repertoire challenging a given genome. As a first step to address this problem, we sought to characterize developmental changes in the expression of PIWI paralogs and piRNAs in a set of species with a conserved repertoire of PIWI paralogs but clear differences in genomic TE landscapes and developmental regimes. Herein, we compare two rodents, mouse and thirteen-lined ground squirrel (*Ictidomys tridecemlineatus*), and the European rabbit (Lagomorpha: *Oryctolagus cuniculus*). The mouse has a typical mammalian genomic TE landscape, dominated by currently active retrotransposons, LINE-1 (L1), and associated SINEs. By contrast, genomic surveys indicate that the last TE insertions in the thirteen-lined ground squirrel genome were deposited between four and five million years ago (Platt and Ray 2012). Thus, the squirrel offers a unique opportunity to study piRNA biogenesis, piRNA diversity and the ping-pong cycle in a system where TE mobilization appears to have ceased. Lastly, rabbit, which exhibits a typical mammalian TE landscape (but distinct from mouse), was included as an outgroup. Specifically, we explored relationships between TE abundance at the genome and transcript level in each species, characterized developmental changes in the expression of PIWI paralogs and piRNAs, determined whether squirrel piRNAs exhibit the signatures of the ping-pong cycle, and explored whether changes in the expression of PIWI paralogs in rabbit and squirrel are

linked to changes in the piRNA repertoires, as is the case with mouse. We found clear differences in PIWI expression and piRNA repertoires when comparing squirrel to rabbit and mouse, which could be linked to the TE shutdown seen in the squirrel. Further, we found evidence of a limited ping-pong response in the squirrel, even in the presence of TE-derived transcripts, suggesting that PIWI-mediated TE silencing is less necessary to maintain genome integrity when the TEs themselves exhibit reduced accumulation rates.

Results

TE accumulation

To understand the evolutionary consequences of the interplay between TEs, piRNAs and PIWIL paralogs, we explored variation in 1) TE activity, 2) the expression and origin of piRNAs, and 3) the expression of PIWI paralogs in rabbit, mouse, and squirrel in a comparative framework. As a first step, we characterized patterns of TE accumulation, diversity, and abundance in the genomes. We assessed the action of TEs by measuring their rates of transcription and of integration into the host genome. As in most mammals, LINEs, SINEs and LTR retrotransposons account for the vast majority of TE insertions in these three species. Close to half of the TE-derived portion of the genome corresponds to LINEs, with a low number of insertions from DNA transposons in all three genomes (Fig. 1A). In contrast, there were clear differences in the relative contribution of SINEs, and LTRs among them: SINEs account for close to half of the TE-derived portion of the genome in the rabbit and LTR retrotransposons contribute a larger fraction in squirrel and mouse relative to the rabbit. These three species also differ in the historical patterns of TE insertion. To reconstruct TE deposition history we classified insertions by subfamilies and estimated the Kimura 2-parameter distance between consensus sequences of each TE family and individual insertions under the assumption that TE copies that are more similar to the corresponding consensus are younger. In the case of the squirrel, these analyses confirm the absence of young TE insertions as reported previously (Fig. 1B) (Platt and Ray 2012). By contrast, in rabbit over 1% of the genome derives from young SINEs, and over 2% of the mouse genome is occupied by young LINE insertions. Here, we define “young” as exhibiting <5% divergence from the relevant consensus sequence.

TE expression

TE expression is expected to be a major determinant of the actual risk a TE poses to the host genome. Further, in the ping-pong model, TE expression is hypothesized to play a major role in shaping piRNA repertoires: transcripts from the most highly transcribed TE subfamilies would be a major source of sense piRNAs, which would then generate the corresponding antisense piRNAs through the ping-pong cycle. Thus, it was critical for our goals to estimate levels of TE expression. We did so by mapping RNA-Seq reads to the libraries of TE consensus sequences for each species. We combined RNA-Seq data from ground squirrel and rabbit testes generated in our study with publicly available data from the mouse. TEs contributed 6%, 4%, and 1% to the mRNA pools in mouse, squirrel, and rabbit, respectively, with SINEs as the most highly expressed TE in all three (Supplemental Fig. 1, Supplementary Online Material). These results were somewhat unexpected for two reasons. First, given the dearth of young TE insertions in the squirrel, we predicted that TE-derived mRNAs would be present at very low levels in this species. Second, because the relative contribution of recent TE insertions to the genome are similar in mouse and rabbit, we did not expect to see major differences in the relative contribution of TEs the corresponding transcriptomes. In addition, we expected that for a given TE type, the relative abundance of recent TE insertions would reflect the relative abundance of the corresponding TE-derived transcripts in the RNA pools. For example, given the small number of young TE insertions in the squirrel, we predicted that TE-derived mRNAs would be present at very low levels. Further, we predicted that the high numbers of young LINE and SINE insertions in the rabbit would be accompanied by a higher transcription of these two types of TEs. We found that, in general, the TE subfamilies with the youngest insertions were the most highly expressed and that just a few subfamilies contributed the majority of RNA (Supplemental Fig. 2, Supplementary Online Material), similar to what was described previously for dog, horse and a microbat (Vandewege et al. 2016). However, we also observed a mismatch between transcript abundance and recent insertions. At the intraspecific level, SINEs in rabbits were expressed at much lower levels than LTRs, but they contribute the most recent insertions (Fig. 1). In mouse, SINEs are also the most expressed TEs but the majority of recent insertions correspond to LINEs. The squirrel represents the most extreme case in our study, as LINEs, LTRs, and SINEs are expressed at similar levels as in the other two species but have not been accumulating.

PIWIL paralog expression

The role of piRNAs in silencing TEs is mediated by the PIWI paralogs, so it was of interest to characterize the expression profiles of the three PIWIL paralogs. To do so we sequenced testis RNA pools from juvenile (0, 2, 8, 10, and 13 days post-birth) and adult squirrels, from juvenile (21 days post-birth) and adult rabbits and incorporated data from juvenile (7 days post-birth) and adult mice for comparative purposes. Our analyses revealed that in all species PIWIL1 is the most highly expressed paralog in adults followed by PIWIL2 (Fig. 2A). In juveniles, this order was reversed, with PIWIL2 expressed at higher levels than PIWIL1. In rabbit and mouse, PIWIL2 was the most expressed paralog in juveniles, as was the case with the p0 and p2 samples from squirrel, but PIWIL4 was the most expressed paralog in the squirrel P8, P10 and P13 samples. Further, there was substantial expression of PIWIL4 in adult squirrel, but expression was not detectable in adult mouse or rabbit. Thus, expression patterns of PIWIL4 constitute the major difference between the squirrel and the other two species, and if the expression of PIWIL4 is ignored, the patterns then become consistent across all three species. We then verified that the adult expression of PIWIL4 is the derived condition in squirrel by comparing to previous data from adult dog and horse testes (Vandewege et al. 2016) (Supplemental Fig. 3, Supplementary Online Material). Thus, our results suggest that squirrel PIWIL4 expression represents a significant outlier relative to other mammals.

piRNAs

As expected, we observed developmental changes in piRNA expression. In squirrel, the P0 and P2 small RNAs libraries exhibited no clear hallmarks associated with PIWI processing of piRNAs, as there were few small RNAs between 24 and 32 nucleotides long, the uridine bias at the 5’ position was absent (Fig. 2B), and almost half of these small RNAs were derived from genic regions (Fig. 2C), not TEs. After P8, we detected the presence of piRNA-like small RNAs with the expected length distribution and nucleotide composition (Fig. 2B and 2C). A slightly higher fraction of these piRNAs mapped to TEs, as expected. This length distribution centered at 27-28 nucleotides persisted into adulthood. This size distribution also distinguishes the adult squirrel from mouse and rabbit, where there is a shift in piRNA size distribution from 27-28 to 29-30 nucleotides associated with the transition to adulthood (Fig. 2B). piRNA mapping patterns showed developmental

differences as well (Fig. 2C). In the three species the percentage of piRNAs that map to genic regions decreases at the transition to adulthood, while the percentage of piRNAs mapping to unannotated portions of the genome increases. The squirrel also stands out in this regard, as there is a decrease in the percentage of piRNAs mapping to TEs in rabbit and mouse that is not clear in the squirrel.

piRNAs and TE dynamics

We then mapped piRNAs and RNASeq data to TE consensus libraries to study how the ping-pong cycle affects TE transcript abundance. As a first step, we examined how ping-pong piRNAs abundance changes in the most expressed L1 among species and development stages (Fig 3). Overall, piRNAs mapped much more densely to L1s in the juvenile stages, except for the squirrel, where piRNA density starts low in the juvenile and stays low in adult. When we compared the expression of TEs and piRNAs among all TEs, once again, squirrel differed from mouse and rabbit. In the squirrel, we found that increases piRNAs mirrored increases in TE expression, with the highest expression of TEs and their associated piRNAs seen in adults (Fig. 4). Further, we did not observe an abundance of ping-pong piRNAs prior to P8 (Fig. 4; Supplemental Fig 4. Supplementary Online Material), consent with the expression of PIWIs (Fig. 1A). In the case of mouse and rabbit, developmental changes in TE expression were inversely related to piRNA abundance in LINEs, LTRs, and rabbit SINEs, but not in mouse SINEs (Fig. 4 B). Interestingly, piRNA pools in squirrel bear a weak 10-nucleotide complementary overlap that is the hallmark of the ping-pong cycle (Fig. 5C). These signatures are evident in juvenile and adult LINE-mapping, and adult SINE-mapping piRNAs in rabbit and mouse, but are not detectable in juvenile SINE-mapping piRNAs in these two species (Fig 5A and B). Taken together, our results suggest that the activity of the ping-pong cycle is substantially reduced among squirrel TEs. Further, a focus on mouse and rabbit would also suggest that the ping-pong cycle responds differently to different TEs throughout development. LINEs appear to be targeted by piRNAs from the early stages of development and decrease, whereas a SINEs appear to be more efficiently targeted in adults but not in juveniles, an observation also reported by Mourier (2011).

Discussion

PIWI proteins and piRNAs have emerged as key components in protecting genomes against TEs, but the interplay among them remains poorly understood. Exploring this interaction in an evolutionary context incorporating a developmental dimension was the central motivation of our study. We focused on comparing the squirrel, mouse, and rabbit because they share a common repertoire of PIWI proteins, but have different TE landscapes. Specifically, we characterized historical patterns of TE accumulation in the corresponding genomes, and compared expression of TEs, piRNAs, and PIWIL paralog expression in from different life stages.

TE expression and accumulation

We predicted a positive relationship between TE expression and the number of recent insertions, but both the rabbit and squirrel yielded data that do not match this prediction. In the case of the squirrel, we found that despite the absence of young insertions in the genome, TEs are still being actively transcribed: We observed relatively high expression of two L1, several SINE, and several LTR subfamilies (Supplemental Fig. 2, Supplementary Online Material). This is consistent with the idea that transcription and successful reintegration are distinct phenomena and that the first does not necessarily lead to the second (Deininger et al. 2003; Kazazian 2004). This is particularly striking in the case of L1s, because the squirrel genome assembly lacks full length, intact L1s, and as a consequence, we expected to see low or no L1 expression. L1 expression is probably maintained by the presence of intact promoters that facilitate transcription, but faulty enzyme products prevent reverse transcription and integration. The presence of active TE transcription in mouse and rabbit was not surprising. However, we found that several mouse SINEs were highly expressed, although they do not appear to be accumulating in its genome. SINEs require the reverse transcriptase and endonuclease machinery encoded by the LINEs to propagate, we speculate that the lack of recent SINE insertions might be due to reduced success by SINEs as they compete for the LINE enzymatic machinery (Boeke 1997; Kroutter et al. 2009). Perhaps the most surprising observation was the overall paucity of TE-derived mRNAs in the rabbit RNA pools, as the number of recent TE insertions in mouse and rabbit are very similar, yet LINE and SINE transcripts are ~ 50 times more abundant in mouse (Supplemental Fig. 1, Supplementary Online Material). Thus, our results raise questions regarding the relationship between TE expression and insertion rates. In fact, our analyses would suggest that insertion success may not be dependent on high levels of TE expression, in line with theoretical predictions from

Lu and Clark (2010). In this regard, it is possible that what constitutes “high” expression may vary among species, and comparisons of TE expression level may not be as informative as is commonly assumed. Or, analogous to mouse, where LINEs and LTR retrotransposons are more highly methylated than SINEs (Molaro et al. 2014), SINEs could be methylated to a larger extent in rabbit testis, thus inhibiting transcription.

PIWIL paralogs

Of the three PIWI paralogs present in rodents, PIWIL1 and PIWIL2 behave as expected, increasing their expression as the animal matures to an adult. In the squirrel, however, the retention of PIWIL4 expression from P8 through adulthood was striking and is not seen in either mouse or rabbit, or any other mammal surveyed. PIWIL4 is only expressed for a few days before and after birth in mice and is expressed in the P21 rabbit but not in the adult. While we cannot make definitive remarks on the functional role of PIWIL4 expression and its relation to piRNA and TE dynamics, the prolonged expression of a PIWI protein focally involved in the silencing of TEs through both transcriptional cleavage and TE methylation is a compelling observation, especially in light of the fact that an L1 apparently has not retrotransposed successfully in the squirrel genome within the last five million years (Platt and Ray 2012). We speculate that the constant expression of PIWIL4 may have affected the squirrel TE landscape by preventing the proliferation of TEs. With no new functional copies being introduced, this may have allowed for all existing L1 elements in the genome to be disabled via mutations in both ORF1 and 2. In future analyses, this aspect might be more thoroughly addressed through genome-wide assessments of methylation, as well as protein visualization and localization experiment assays at different development stages.

piRNAs and TEs

Our analysis suggests that the interplay between piRNAs and TEs is complex, changes throughout development, and has some lineage-specific features. In our study, both mouse and rabbit have an active TE landscape and their piRNA repertoires bear the hallmarks of the ping-pong cycle. In these two species, with the exception of mouse SINEs, there is an inverse relationship between developmental changes in TE abundance and changes in the abundance of the corresponding piRNA pairs (Fig 4). This observation matches the prediction of the ping-

pong model. Pre-pachytene piRNAs are known to be biased towards TEs relative to pachytene piRNAs: a pattern observed in both the mouse and rabbit, but largely absent from the squirrel, which suggests that the ping-pong cycle is not currently playing a strong role in repressing TE expression. In fact, multiple lines of evidence suggest the absence of an active response to TEs in the squirrel through the ping-pong cycle, even though TEs are being transcribed. In this species, we found that increases in TE expression were associated with increases in the number of piRNA ping-pong pairs, with the highest expression of TEs and the associated piRNAs in adults (Fig. 4). Furthermore, the overlaps among these ping-pong pairs have a relatively weak peak at 10 base pairs (Fig. 5), typical of the ping-pong cycle, as seen in the case of adult mouse and rabbit.

In mammals that are currently accumulating TE insertions, there are developmental changes in how the piRNAs map to TEs. Gainetdinov et al. (2017) compared fetal and adult human testis and suggest that there are two complementary modes of TE control, where pre-pachytene piRNAs mostly target recently active TEs, and pachytene piRNAs target older insertions. Our study also identifies developmental differences in how piRNAs map to the most highly expressed LINE in juvenile and adult mouse and rabbit, where piRNAs in juveniles but not adults in the corresponding comparison in squirrel (Fig. 3). In addition, there also changes observed in mouse and in rabbit, the two species with active TEs and traces of an active ping-pong cycle. Whereas LINE activity appears to elicit an active ping-pong response starting at the juvenile stage, SINEs are apparently not targeted by the ping-pong cycle until the adult stage (Fig 5).

Conclusions

Taken together, our results raise interesting questions regarding the interplay between TEs, piRNAs and PIWIL paralogs. To begin with, our study indicates that TE transcript abundance is not a good predictor of TE accumulation: TEs appear to be expressed much more in mouse than in rabbit, but the two genomes appear to be accumulating TEs at a similar rate, and within each of them, the most expressed TEs were not the ones generating the most insertions. In addition, we found that in these two species that harbor active TEs in the genomes, ping-pong piRNAs appear to be reducing the abundance of TE transcripts. In the case of the squirrel, we effectively asked what PIWI/piRNAs do when TEs are not actively mobilizing. Contrary to our expectations,

we found that TEs are being actively transcribed in the squirrel even though they do not accumulate in the genome and that these TEs appear not to be actively targeted by ping-pong piRNAs. Unlike the profiles observed in other mammals, in the squirrel, the presence of piRNAs mapping to a given TE did not lead to any reduction of TE transcripts and bears no traces of the ping-pong cycle. We hypothesize that this derives from the passive processing of TE transcripts into piRNAs in the squirrel, which sharply contrast with the PIWI/piRNA associated active reduction of TE transcripts in the rabbit and mouse and other mammals screened to date. In fact, the piRNA repertoires of the squirrel suggest the absence of an active ping-pong cycle in this species, a unique observation. The expression of PIWIL4 in adult testis also distinguishes the squirrel from other mammals and would suggest that methylation of TE-like sequences might have played a role in the TE shutdown experienced by the squirrel. The extended expression of PIWIL4 in squirrel should make this species an attractive model to explore the functional role of this PIWIL paralog and warrants further assays.

Methods

Sample acquisition and RNA sequencing

All animal procedures conformed with federal and institutional guidelines for humane care and use and were pre-approved by the respective Institutional Animal Care and Use Committees. Thirteen-lined ground squirrels were reared in the captive breeding colony at the University of Wisconsin, Oshkosh (Merriman et al. 2012). For testis collection, euthanasia by cervical transection was conducted on neonates (P0), post-natal juveniles (P2, P8, P10, P13), and adults. For the rabbits, we sampled one juvenile (P21) and one adult (P > 90) male rabbit. For each specimen, a cross section of testis was snap frozen in liquid nitrogen immediately following castration and stored at -80°C prior to RNA isolation. We isolated total RNA using Trizol ® (Invitrogen, USA) according to the manufacturer's specifications. Small RNA libraries were prepared using the Illumina TruSeq small RNA kit © and 1 x 50 bp reads were sequenced on the Illumina Hi-Seq2000 platform. Directional RNASeq libraries were prepped using the Illumina TruSeq v2 kit and 2x100 bp reads were also sequenced on a HiSeq2000. All reads are available under the BioProject PRJNA528042. P7 mouse mRNA (SRR3659160), small RNAs (SRR3659150) adult mRNA (SRR765631) and mRNA (SRR772033) libraries were collected from the short-read archive (SRA).

TE annotation

To visualize the TE differences between the squirrel, rabbit and mouse genomes, we masked the ground squirrel (SpeTri2.0), rabbit (OryCun2.0.73) and mouse (GRCm38) genome using RepeatMasker 4.0.5 '-species *Ictidomys*', '-species *Oryctolagus*', and '-species *Mus*', respectively. To estimate genetic distances, we used a modified calcDivergenceFromAlign.pl script included with RepeatMasker to calculate Kimura two-parameter (Kimura 1980) distances between each insertion and its respective consensus sequence. The option –noCpG was invoked to exclude highly mutable CpG sites from distance calculations. Redundant TE annotations were first identified using clusterBed (Quinlan and Hall 2010) and the longest annotation with the lowest K2P was selected as the most probable.

Genomic mapping

Prior to small RNA mapping, we clipped barcodes, removed reads that had bases with Phred quality score <25, and removed identical reads using modules in the fastx toolkit. We also removed low complexity small RNA sequences by zipping the sequence and removing sequences that compressed by more than 75%. We mapped piRNA sequences 24-32 bases long to the respective genomes using Bowtie 1.2 (Langmead et al. 2009) allowing one mismatch in the alignment and only allowed piRNAs to map to the 'best' locus (bowtie -y -v 1 -best). Ensemble annotations were used to define UTR, exon and intron boundaries. Each piRNA was annotated as TE-derived, genic, or unannotated.

TE expression and piRNA dynamics

To estimate expression profiles through testis development, we mapped RNASeq reads to a combined library of Ensembl predicted gene CDS sequences and TE consensus sequences using RSEM 1.3.1 (Li and Dewey 2011). If there were multiple predictions for a given gene, only the longest sequence was used to estimate expression. We used default RSEM parameters and measured relative as transcripts per million (TPM).

Further, we examined piRNAs exhibiting evidence of the ping-pong cycle for TE consensus sequences. We mapped small RNA reads to TE consensus sequences using bowtie and allowed piRNAs to map to all positions allowing three mismatches per alignment (bowtie -a -v 3). We filtered for piRNAs that exhibited the ping-pong signature: small RNAs with a uridine in the first position of a primary piRNA, the primary piRNA overlapped with a complementary piRNA by 10 bp and the complementary (secondary) piRNA had an adenine in the 10th position (Vandewege et al. 2016). To compare piRNA abundance among species and developmental stages, we created a metric that corrected for piRNAs that mapped to multiple positions, library size, and consensus sequence lengths; we called this metric corrected piRNAs per kilobase million (cPPKM). Specifically, we scored each piRNA by dividing one by the number of different mapped positions. For each consensus sequence we summed the scores of the mapped piRNAs, divided by the score by the length of the consensus sequence (in thousands) and divided this number by the total number of piRNAs that mapped to the consensus sequence library (in millions).

To observe patterns between TE expression and ping-pong piRNA abundance, we summed TE expression values for each major TE subclass (LINEs, SINEs, and LTRs) and corresponding cPPKM values. We did not include DNA transposons in our analyses because these TEs were uncommon in the genome and had almost no expression and very few mapped piRNAs.

Acknowledgments

We acknowledge the support from the National Science Foundation (DEB-1355176, RoL 1838283). Additional support was provided by the College of Agriculture and Life Sciences at Mississippi State University and the College of Arts and Sciences at Texas Tech University. In addition, we would like to thank the Texas Tech High-Performance Computing Center and the Mississippi State University Institute for Genomics Biocomputing and Biotechnology for providing the computational resources.

References

Aravin AA, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, Morris P, Brownstein MJ, Kuramochi-Miyagawa S, Nakano T, et al. 2006. A novel class of small RNAs bind to MILI protein in mouse testes. Nature 442:203–207.

Aravin AA, Hannon GJ, Brennecke J. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science 318:761–764.

Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF, Bestor T, Hannon GJ. 2008. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. Mol. Cell 31:785–799.

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. 2007. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science 316:744–747.

Boeke JD. 1997. LINEs and Alus — the polyA connection. Nat. Genet. 16:6–7.

Brennecke J, Aravin A a., Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell 128:1089–1103.

Carmell MA, Girard A, van de Kant HJG, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ. 2007. MIWI2 Is essential for spermatogenesis and repression of transposons in the mouse male germline. Dev. Cell 12:503–514.

Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Pääbo S, et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437:88–93.

Deininger PL, Moran J V, Batzer MA, Kazazian HH. 2003. Mobile elements and mammalian genome evolution. Curr. Opin. Genet. Dev. 13:651–658.

Ernst C, Odom DT, Kutter C. 2017. The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. Nat. Commun. 8:1411.

De Fazio S, Bartonicek N, Di Giacomo M, Abreu-Goodger C, Sankar A, Funaya C, Antony C, Moreira PN, Enright AJ, O'Carroll D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. Nature 480:259–263.

Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J, et al. 2017. Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res. 27:1384–1394.

Gainetdinov I, Skvortsova Y, Kondratieva S, Funikov S, Azhikina T. 2017. Two modes of targeting transposable elements by piRNA pathway in human testis. RNA 23:1614–1625.

Girard A, Sachidanandam R, Hannon GJ, Carmell MA. 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature 442:199–202.

Gou L-T, Dai P, Yang J-H, Xue Y, Hu Y-P, Zhou Y, Kang J-Y, Wang X, Li H, Hua M-M, et al. 2014. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. Cell Res. 24:680–700.

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov D V., Blaser H, Raz E, Moens CB, et al. 2007. A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in zebrafish. Cell 129:69–82.

Kazazian HH. 2004. Mobile elements: drivers of genome evolution. Science 303:1626–1632.

Kerner P, Degnan SM, Marchand L, Degnan BM, Vervoort M. 2011. Evolution of RNA-binding proteins in

animals: insights from genome-wide analysis in the sponge Amphimedon queenslandica. Mol. Biol. Evol. 28:2289–2303.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:111–120.

de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7:e1002384.

Kroutter EN, Belancio VP, Wagstaff BJ, Roy-Engel AM. 2009. The RNA Polymerase Dictates ORF1 Requirement and Timing of LINE and SINE Retrotransposition.Malik HS, editor. PLoS Genet. 5:e1000458.

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:R25.

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE. 2006. Characterization of the piRNA complex from rat testes. Science 313:363–367.

Lewis SH, Salmela H, Obbard DJ. Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323.

Li XZZ, Roy CKK, Dong X, Bolcun-Filas E, Wang J, Han BWW, Xu J, Moore MJJ, Schimenti JCC, Weng Z, et al. 2013. An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. Mol. Cell 50:67–81.

Lu J, Clark AG. 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. Genome Res. 20:212–227.

Malone CD, Hannon GJ. 2009. Small RNAs as guardians of the genome. Cell 136:656–668.

Molaro A, Falciatori I, Hodges E, Aravin A a., Marran K, Rafii S, Richard McCombie W, Smith AD, Hannon GJ. 2014. Two waves of de novo methylation during mouse germ cell development. Genes Dev. 28:1544–1549.

Mourier T. 2011. Retrotransposon-centered analysis of piRNA targeting shows a shift from active to passive retrotransposon transcription in developing mouse testes. BMC Genomics 12:440.

O'Donnell KA, Boeke JD. 2007. Mighty Piwis defend the germline against genome intruders. Cell 129:37–44.

Platt RN, Ray DA. 2012. A non-LTR retroelement extinction in Spermophilus tridecemlineatus. Gene 500:47–53.

Platt RN, Vandewege MW, Ray DA. 2018. Mammalian transposable elements and their impacts on genome evolution. Chromosom. Res. 26.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Sánchez Alvarado A. 2005. SMEDWI-2 Is a PIWI-Like

Protein That Regulates Planarian Stem Cells. Science (80-.). 310:1327–1330.

Siomi MC, Sato K, Pezic D, Aravin A a. 2011. PIWI-interacting small RNAs: the vanguard of genome defence. Nat. Rev. Mol. Cell Biol. 12:246–258.

Vandewege MW, Platt RN, Ray DA, Hoffmann FG. 2016. Transposable element targeting by piRNAs in Laurasiatherians with distinct transposable element histories. Genome Biol. Evol. 8:1327–1337.

Figures:

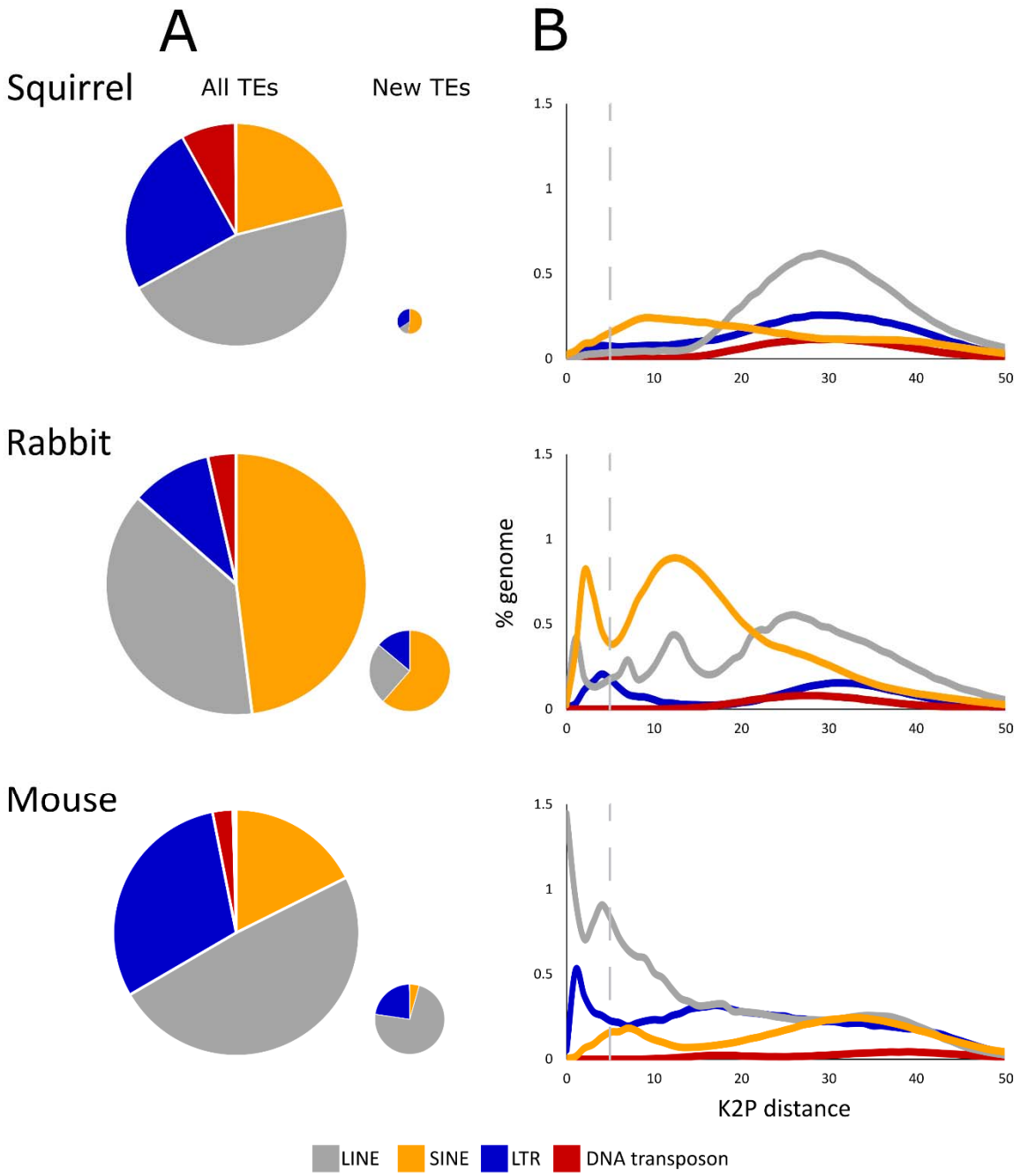

Figure 1. A) Relative contribution of TEs to the corresponding genome. The small pie charts document the contribution of the more recent TE insertions, with the size of the pie chart proportional to the relative contribution of recent insertions to each genome. B) Historical patterns of accumulation of the major TE categories inferred by calculating the Kimura 2-parameter distance between individual insertions and the corresponding consensus. Relatively young (<5% diverged) insertions to the left of the vertical line with lower genetic distances were deposited more recently. The insertions to the left of the dashed vertical lines were considered for the small pie charts.

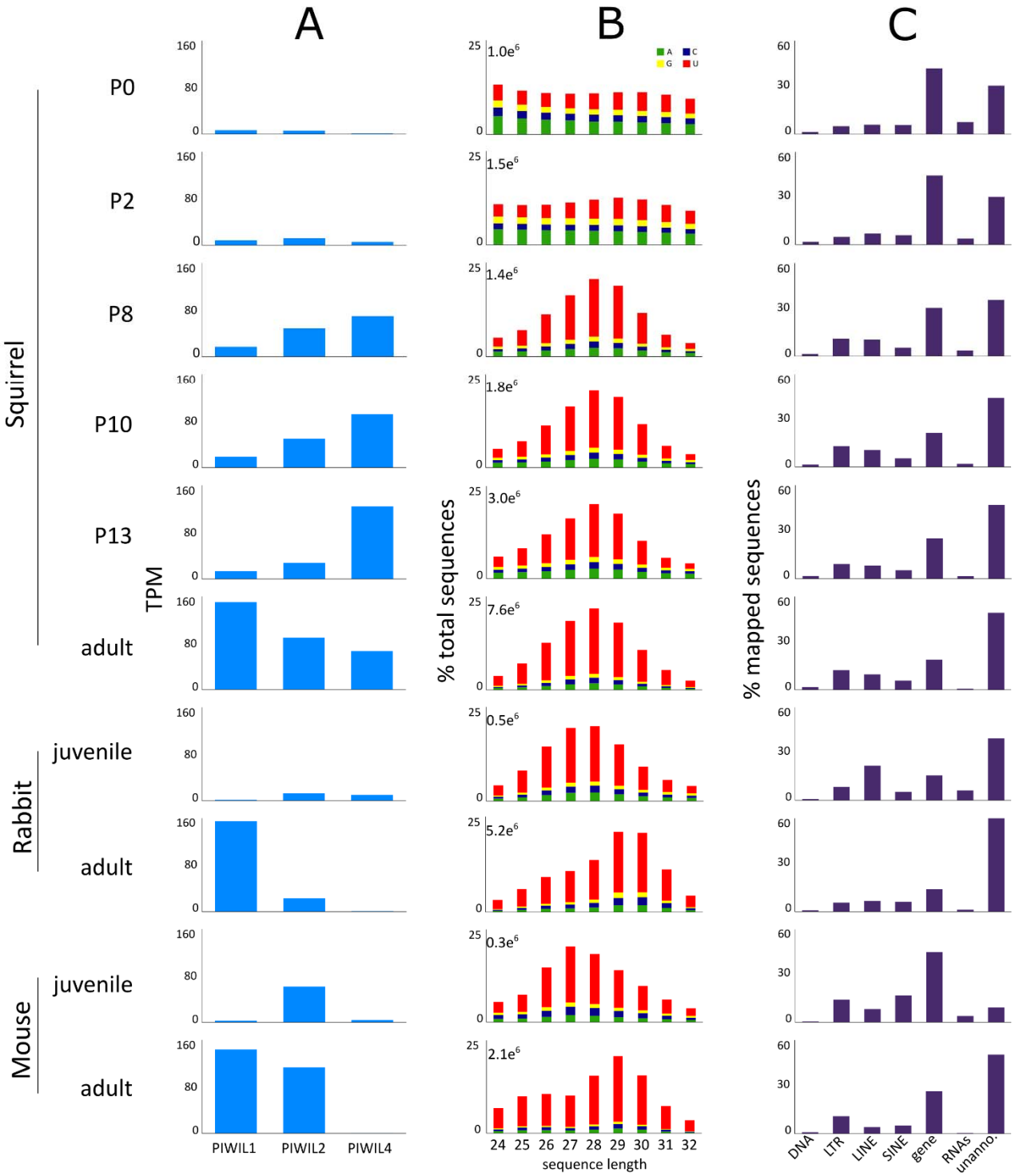

Figure 2. A) Expression of the three PIWI proteins among species and developmental stages. Expression is measured as transcripts per million (TPM). B) Length distribution and first base composition of unique small RNAs presented as a proportion of reads between 24 and 32 bases long. Numbers on the upper left denote the number of reads considered. C) Proportion of genomic small RNA annotations. Genes includes UTRs, introns and exons. If a small RNA mapped to a TE within an intron, we annotated the small RNA as TE-derived.

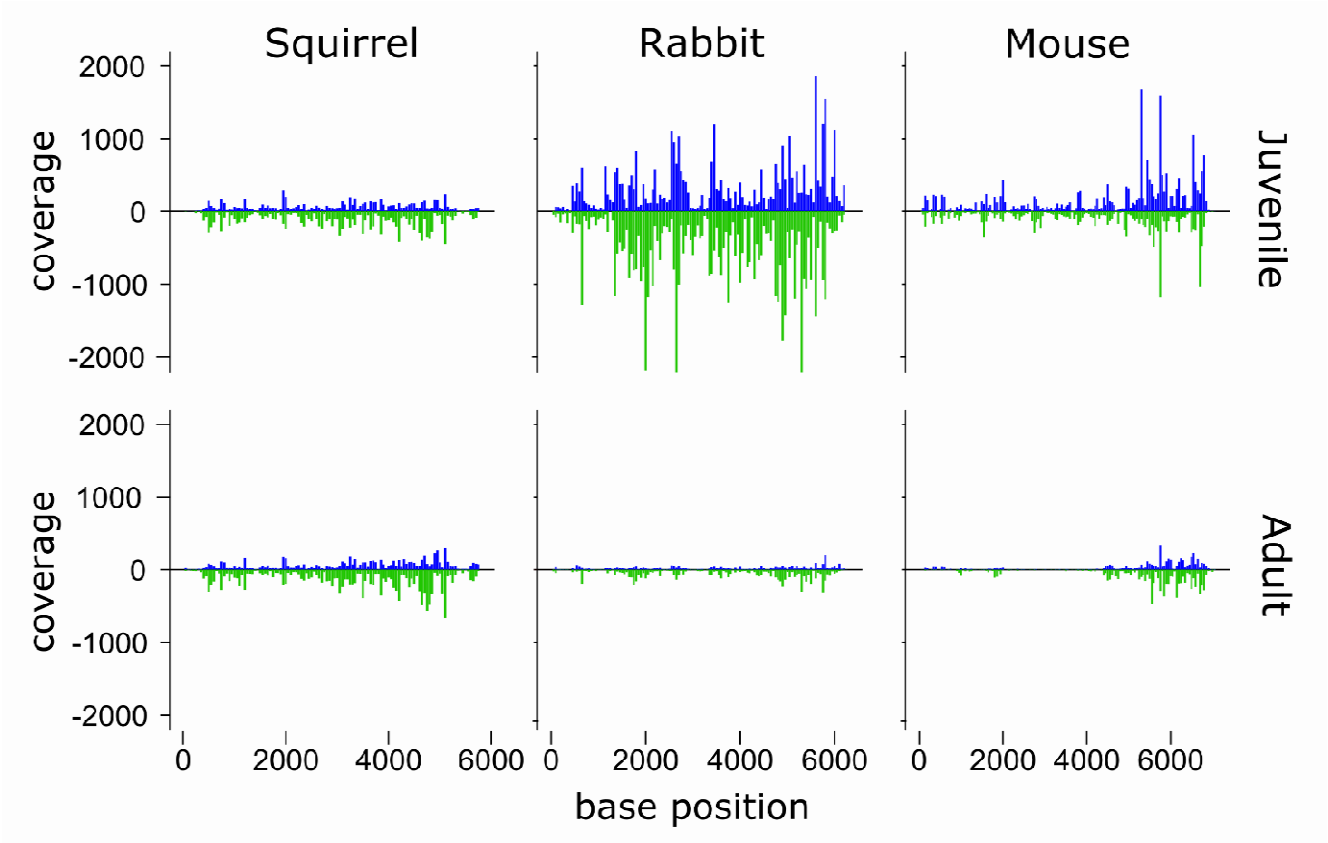

Figure 3. Raw ping-pong piRNA coverage along the most expressed L1 between juvenile and adult development stages, where blue is sense coverage and green is anti-sense coverage for A) squirrels, B) rabbits, C) mice. Individual bars represent the averages of 50 bp windows. Juvenile piRNA coverages are on top and adults are below. Development stage P10 represented the juvenile squirrel; P21 the juvenile rabbit; and P7 the juvenile mouse.

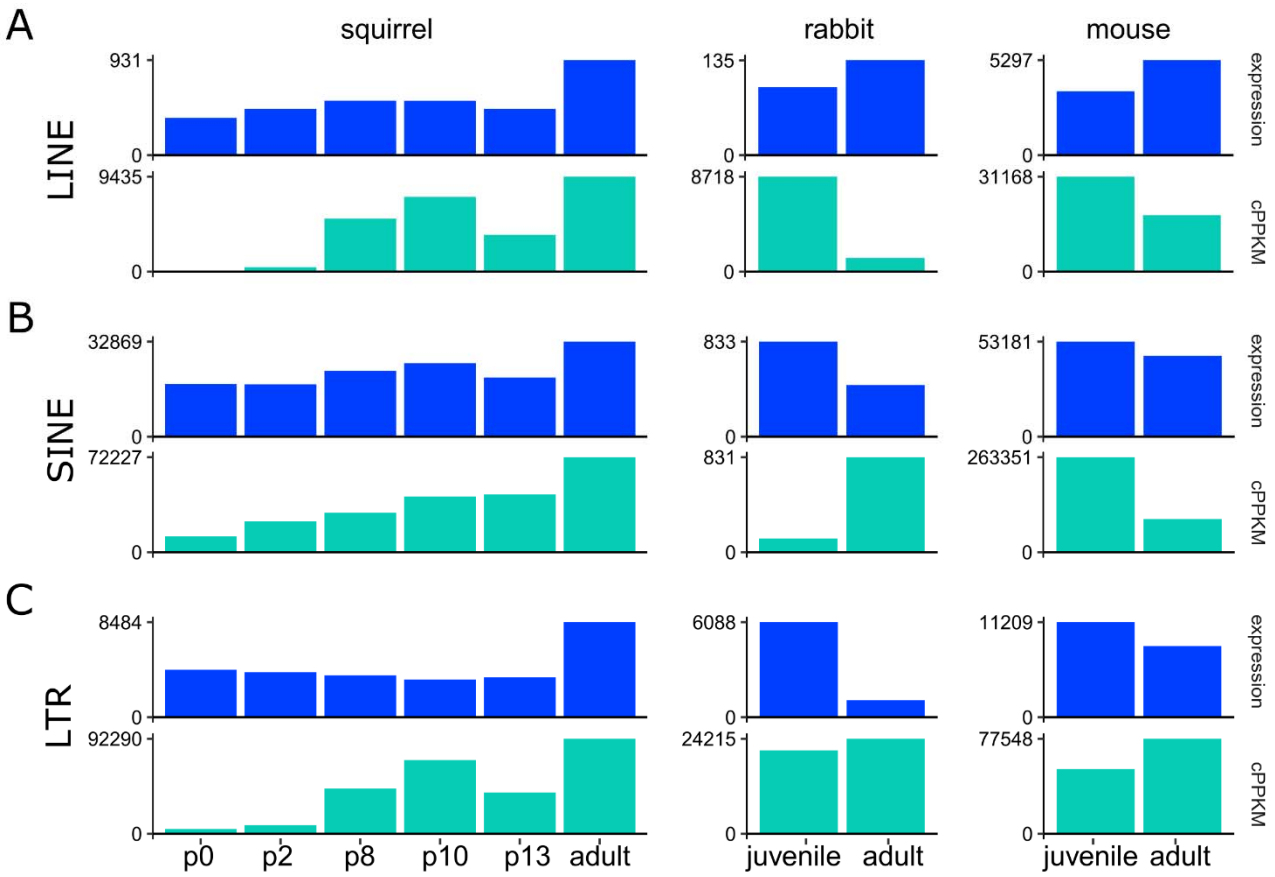

Figure 4. Developmental changes in expression of TEs in blue, and piRNAs in green for A) LINEs, B) SINEs, and C) LTRs. For each TE group, the values TE expression (in TPM) and piRNA abundance (cPPKM) were summed from all consensus sequences. Development stage P21 represented the juvenile rabbit; and P7 represented the juvenile mouse.

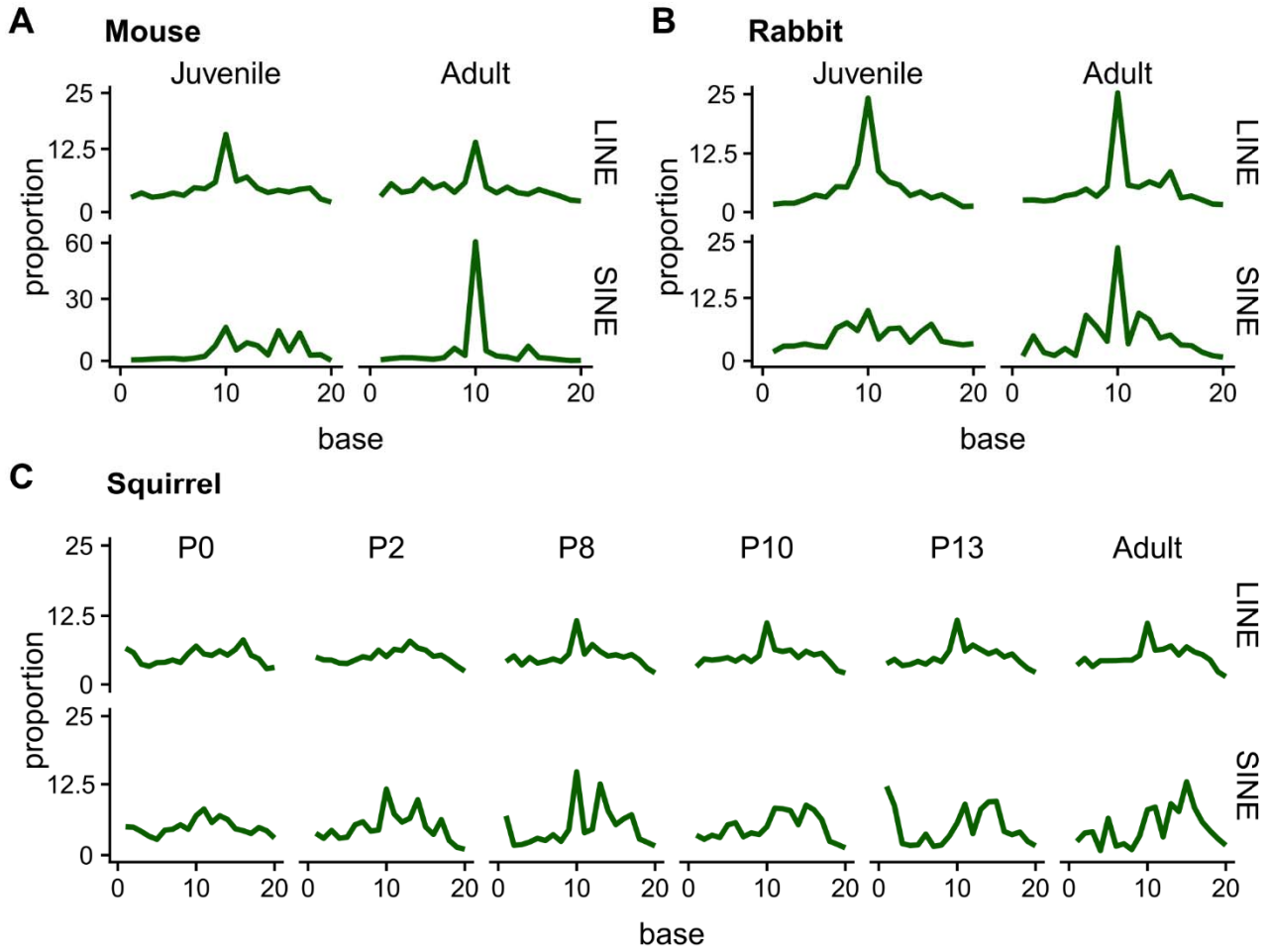

Figure 5. The proportion of base pairs that overlap among complementary small RNAs mapped to the most expressed LINE or SINE consensus sequence in each species per development stage for the A) mouse B) rabbit and C) squirrel. Development stage P21 represented the juvenile rabbit; and P7 represented the juvenile mouse.