



Transposable elements drive the evolution of metazoan zinc finger genes

Jonathan N. Wells, Ni-Chen Chang, John McCormick, et al.

Genome Res. 2023 33: 1325-1339 originally published online September 15, 2023

Access the most recent version at doi:[10.1101/gr.277966.123](https://doi.org/10.1101/gr.277966.123)

References This article cites 118 articles, 23 of which can be accessed free at:
<http://genome.cshlp.org/content/33/8/1325.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Transposable elements drive the evolution of metazoan zinc finger genes

Jonathan N. Wells,¹ Ni-Chen Chang,^{1,4} John McCormick,^{1,4} Caitlyn Coleman,² Nathalie Ramos,^{1,3} Bozhou Jin,¹ and Cédric Feschotte¹

¹Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14850, USA; ²Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, Florida 33620, USA; ³Department of Genetics and Genomic Sciences, Center for Transformative Disease Modeling, Tisch Cancer Institute, Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA

Cys2-His2 zinc finger genes (ZNFs) form the largest family of transcription factors in metazoans. ZNF evolution is highly dynamic and characterized by the rapid expansion and contraction of numerous subfamilies across the animal phylogeny. The forces and mechanisms underlying rapid ZNF evolution remain poorly understood, but there is growing evidence that, in tetrapods, the targeting and repression of lineage-specific transposable elements (TEs) plays a critical role in the evolution of the Krüppel-associated box ZNF (KZNF) subfamily. Currently, it is unknown whether this function and coevolutionary relationship is unique to KZNFs or is a broader feature of metazoan ZNFs. Here, we present evidence that genomic conflict with TEs has been a central driver of the diversification of ZNFs in animals. Sampling from 3221 genome assemblies, we show that the copy number of retroelements correlates with that of ZNFs across at least 750 million years of metazoan evolution. Using computational predictions, we show that ZNFs preferentially bind TEs in diverse animal species. We further investigate the largest ZNF subfamily found in cyprinid fish, which is characterized by a conserved sequence we dubbed the fish N-terminal zinc finger-associated (FiNZ) domain. Zebrafish possess approximately 700 FiNZ-ZNFs, many of which are evolving adaptively under positive selection. Like mammalian KZNFs, most zebrafish FiNZ-ZNFs are expressed at the onset of zygotic genome activation, and blocking their translation using morpholinos during early embryogenesis results in repression of transcriptionally active TEs. Together, these data suggest that ZNF diversification has been intimately connected to TE expansion throughout animal evolution.

[Supplemental material is available for this article.]

The Cys2-His2 zinc finger (ZNF) is a short nucleic acid binding domain widespread in eukaryotes. ZNFs are often arranged in tandemly repeated arrays, resulting in proteins with versatile nucleic acid binding ability (Wolfe et al. 2000; Klug 2010; Najafabadi et al. 2017). In metazoans, tandem-repeat ZNF genes have undergone lineage-specific expansions in gene copy number, as well as an increase in the diversity of DNA sequences that they recognize (Emerson and Thomas 2009; Najafabadi et al. 2017; Heger et al. 2020). Many ZNF proteins contain accessory domains that modulate chromatin states and transcription activity, such as the repressive Krüppel-associated box (KRAB) (Bellefroid et al. 1991). Thus, ZNFs represent the largest and most dynamically evolving transcription factor family in metazoans.

The abundance and diversity of C2H2 ZNF genes are a hallmark feature of metazoan genomes (Najafabadi et al. 2017; Heger et al. 2020), and in species whose genomes have been sequenced, ZNF gene copy number spans three orders of magnitude, from fewer than ten in many nematodes to hundreds, if not thousands, in some vertebrates, arthropods, and mollusks (Goodstadt et al. 2007; Albertin et al. 2015; Panfilio et al. 2019). Furthermore, they are frequently lineage-specific; for example, a survey of three great ape species and rhesus macaque found that ~10% (213 out of 2253)

of predicted KRAB-ZNFs (KZNFs) were species-specific, including seven from humans and 145 from orangutans (Nowick et al. 2011). This rapid evolution contrasts with most other transcription factors, which are deeply conserved across metazoans (Degnan et al. 2009).

However, the same qualities that differentiate ZNFs from other transcription factors make them challenging to study, and consequently, there have been few attempts to produce a unified theory explaining their prevalence and recurrent expansions (Emerson and Thomas 2009; Fedotova et al. 2017). Early hints at the function of KZNFs came from work showing that deletion of the genes *TRIM28* and *SETDB1*, the H3K9 methyltransferase, led to increases in endogenous retrovirus expression in mice (Matsui et al. 2010; Rowe et al. 2010). Because *TRIM28* was known to interact with the KRAB domain, this suggested that KZNFs may be involved in regulating retroelements (Thomas and Schneider 2011).

Transposable elements (TEs) are selfish genetic elements that replicate independently within their host genomes and, in most eukaryotes, comprise between 5% and 85% of the genome. Uncontrolled TE proliferation has deleterious consequences, ranging from insertional mutagenesis and genomic instability caused by ectopic recombination between TE insertions to dysregulation of gene expression stemming from the fact that many TEs carry their own promoters and regulatory sequences. Consequently, metazoans have evolved a variety of defenses to control their

⁴These authors contributed equally to this work.

Corresponding authors: jnw72@cornell.edu, cf458@cornell.edu

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.277966.123>. Freely available online through the *Genome Research* Open Access option.

© 2023 Wells et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

spread, most notably the piRNA system (Aravin et al. 2007; Czech and Hannon 2016) and the previously mentioned KZNFs (Ecco et al. 2017; Bruno et al. 2019). A hallmark of both of these defense systems is their ability to adapt to changing populations of invading TEs, which leads to selection pressure on the TEs to evade these defenses, thereby creating arms race dynamics (McLaughlin and Malik 2017).

There is growing evidence that KZNFs are engaged in such arm races with TEs (Jacobs et al. 2014; Fernandes et al. 2018; Bruno et al. 2019). KZNFs form the largest ZNF gene subfamily found in Sarcopterygii, namely, tetrapods and lobe-finned fish (Bellefroid et al. 1991; Imbeault et al. 2017). The extensive variation in KZNF copy number across species has long been appreciated (Bellefroid et al. 1991; Huntley et al. 2006), but the evolutionary forces driving this variation remained elusive until a breakthrough study showing that the number of ZNF domains in a given genome is positively correlated with retroelement copy number across a small but diverse sample of vertebrates, suggesting a coevolutionary relationship between the two (Thomas and Schneider 2011). This was bolstered by subsequent ChIP-seq experiments in humans and mice mapping the genome-wide binding of hundreds of KZNFs, which revealed that most target specific TE families (Imbeault et al. 2017; Wolf et al. 2020). Furthermore, KZNF knockouts in mice and humans lead to up-regulation of TE expression (Wolf et al. 2020; Haring et al. 2021). Mechanistic studies showed that KZNF repression of TEs is mediated via their KRAB domain, which interacts with the KAP1/TRIM28 corepressor to recruit SETDB1, among several other chromatin silencing factors (Wolf and Goff 2009; Matsui et al. 2010; Rowe et al. 2010; Ecco et al. 2017); in doing so, most KZNFs nucleate the formation of heterochromatin at their target TE loci. Together, these findings support the idea that TE proliferation drives the diversification of KZNFs.

Much less is known about the factors influencing the evolution of other ZNF families. Here, we investigate the hypothesis that interaction with TEs is a driving force underlying the molecular evolution and functional diversification of ZNF families throughout metazoans. Using a combination of comparative phylogenetics, evolutionary analyses, and in silico predictions of ZNF binding specificity, we explore the coevolutionary relationship between TEs and ZNFs in animals. We then turn our focus to a large family of ZNFs present in zebrafish, which are defined by their association with a conserved protein sequence we have termed the fish N-terminal zinc finger-associated domain (FiNZ). FiNZ-ZNF (FZNF) genes are unique to cyprinid fish and have evolved in parallel to the KZNFs found in tetrapods. Using refined gene annotations in cyprinids, we test whether these novel genes are also evolving under positive selection, consistent with ongoing arm races with TEs. Finally, by simultaneously targeting more than 400 FZNFs for knockdown during zebrafish embryogenesis, we ask whether FZNFs are capable of silencing TE expression.

Results

Annotation of ZNFs and TEs in metazoan genomes

We sought to estimate the number of ZNF genes in a large set of publicly available metazoan genome assemblies. Using open reading frames (ORFs) as a proxy for genes, we searched 3221 metazoan genome assemblies for ORFs containing ZNF domains. As many proteins contain small numbers of standalone ZNF domains (Supplemental Fig. 1), we set a threshold of at least five ZNF do-

main for inclusion in our data set in order to focus on those proteins with tandemly repeated ZNF arrays. This search revealed extensive variation across species, from fewer than 10 ORFs containing at least five ZNFs in most nematode worms to upward of a thousand in many vertebrates, mollusks, and arthropods (Fig. 1A; Supplemental Data 1, 2). Because estimates of ZNFs could be biased by genome assembly quality, we set a minimum scaffold N50 of 50 kb to filter out low-quality assemblies. To ensure that this threshold was appropriate, we calculated the correlation between our ZNF ORFs counts and scaffold N50 length (Supplemental Fig. 2A). This revealed a significant, but weak correlation of $\rho=0.1$, reassuring us that assembly quality was not unduly biasing our ORF counts.

In addition to automated counts, we performed more thorough gene annotations in cyprinid fish (Supplemental Data 2). Cyprinids, which include the widely used model organism *Danio rerio* along with many economically and culturally important carp species, are one of the most species-rich vertebrate families, with more than 1500 extant members (<https://www.calacademy.org/scientists/projects/eschmeyers-catalog-of-fishes>). They possess a large family of ZNF genes characterized by the presence of the FiNZ domain, a short, 28-amino-acid N-terminal sequence (Supplemental Fig. 3A). FiNZ has no significant sequence similarity to known protein domains but is highly conserved across cyprinid ZNFs and is enriched in negatively charged glutamic acid residues, a feature it shares with the KRAB domain (Bonferroni-corrected P -value < 0.00125 for both, relative to sequences in the SwissProt51 database) (Vacic et al. 2007). The canonical structure of FZNF genes resembles that of KZNFs, with the FiNZ domain being encoded by a single exon upstream of a second exon consisting of an array of ZNF domains (Imbeault et al. 2017).

We reannotated putative FZNF genes using a combination of BLAST (Altschul et al. 1990), HMMER (Eddy 2011), and AUGUSTUS-PPX (Keller et al. 2011; Hoff and Stanke 2018). In zebrafish, this approach yielded a total of 684 FZNF genes, a substantial increase from the number annotated in the preexisting RefSeq and Ensembl gene sets (Supplemental Fig. 3B). Almost all FZNF genes in zebrafish are located on the long arm of Chromosome 4 (Chr 4q), a region previously noted for its enrichment in TEs, ZNFs, and immunity genes, as well as a wide variety of RNA-encoding loci (Howe et al. 2013, 2016). This physical colocalization of FZNFs superficially resembles that of KZNF genes, which form large clusters of tandemly repeated genes in mammals (Tadepally et al. 2008).

Across the cyprinid family, the copy number of predicted FZNF genes varies considerably (Fig. 1B). To validate these copy number estimates, we used raw reads from four *Danio* species and compared the read coverage over FZNF genes to coverage over benchmarking universal single-copy orthologs (BUSCO) genes. In three out of four cases, there was good agreement between the two, but in the case of *Danio albolineatus*, the median read depth over FZNF genes was $\sim 50\%$ greater than that of BUSCO genes, indicating that annotations provide a lower bound on copy number (Supplemental Fig. 4A). Finally, to assess whether automated ZNF ORF counts are good proxies for the number of annotated ZNF genes in a species, we compared ORF counts to those of our reannotated FZNFs, as well as independent estimates of KZNF copy number in tetrapods (Imbeault et al. 2017). In both cases, we found good agreement between the simpler ORF counts and higher-quality estimates based on gene annotations (respectively, Spearman's $\rho=0.91$, P -value $= 2.6 \times 10^{-59}$; $\rho=0.97$, P -value $= 1.7 \times 10^{-5}$) (Supplemental Fig. 4B,C).

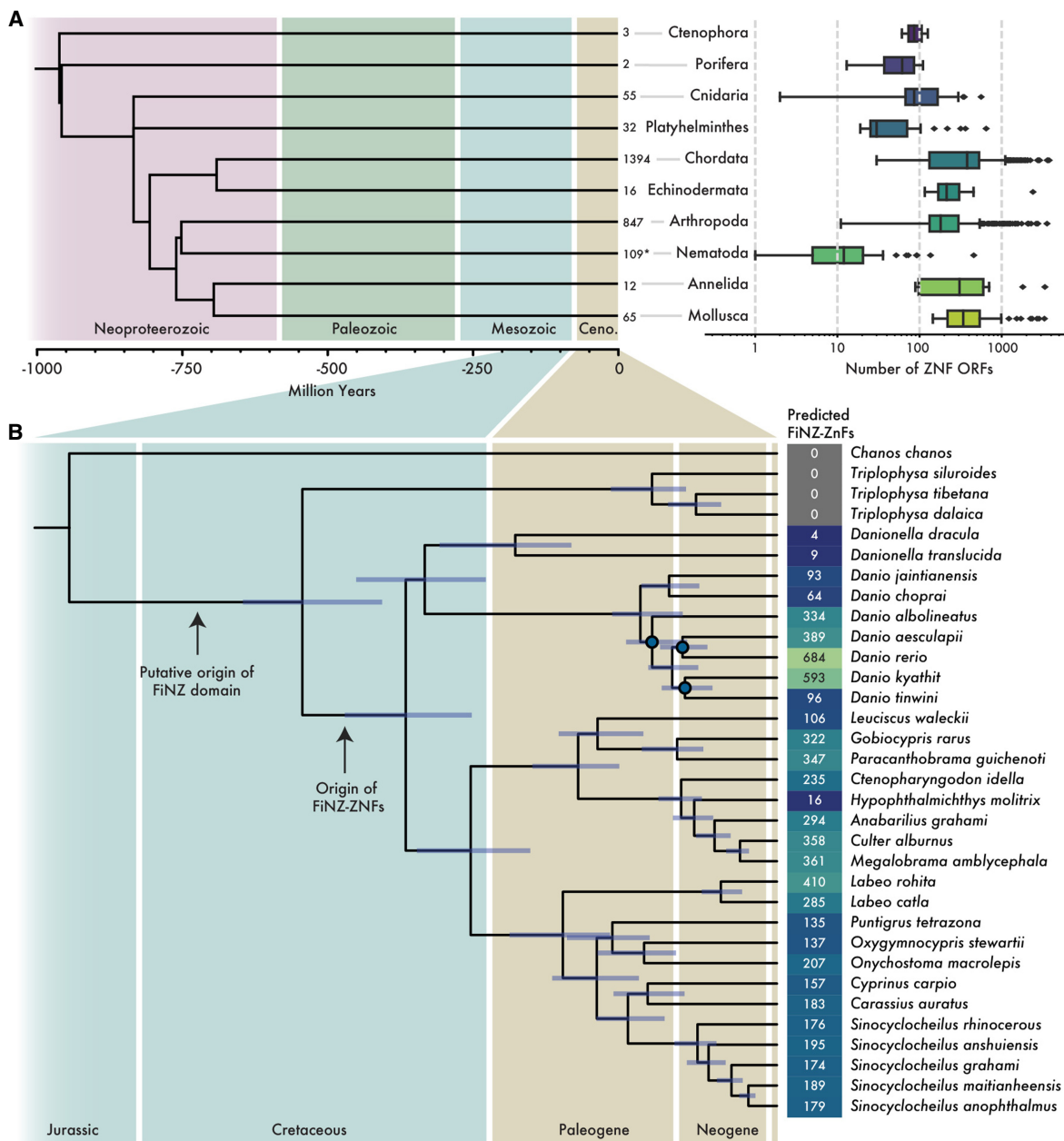


Figure 1. Annotation of ZNF ORFs and genes in metazoan genome assemblies. (A) Boxplots showing distributions of ZNF ORF counts per species in each major taxonomic phylum. *Adjacent* to boxplots are the number of representative species per phylum. (*) Number by nematodes excludes those with zero tandem repeat ZNF ORFs (defined as containing at least five repeats). Phylogeny was acquired from TimeTree5 (Kumar et al. 2022). (B) Maximum likelihood phylogenetic tree of Cypriniformes species with high-quality genome assemblies. Numbers represent counts of FZNF genes using improved gene annotations. Nodes marked in blue have lower than 95/95 support for Shimodaira–Hasegawa approximate likelihood ratio test/ultra-fast bootstraps, respectively.

Correlation between ZNFs and retroelements

Early evidence for a coevolutionary relationship between TEs and ZNFs was reported by Thomas and Schneider (2011), who observed a strong correlation between ZNF and retroelement copy number in a sample of 26 vertebrate species. We sought to reproduce and expand on these results, making use of the greatly increased number of genome assemblies available today. Using the same process as for ZNF ORFs, we searched for reverse transcriptase and RNase H domains, using these as proxy for the retroelement copy number

(Supplemental Data 1, 2). To address the issue of autocorrelation produced by the nonindependence of phylogenetically related species (Felsenstein 1985), we selected a single species per taxonomic family; because ZNF and TE turnover is rapid, representative species from different families have largely independent complements of both (Liu et al. 2014; Imbeault et al. 2017).

From the resulting set of 828 genomes, we examined the correlation between the estimated number of ZNFs and retroelement ORFs. This analysis revealed a significant positive correlation across all metazoans, as well as significant positive correlations

across major phyla with at least 15 representative species, with the exception of nematode worms (Fig. 2A). Within smaller taxonomic groups, this relationship holds in most cases tested (Supplemental Data 1), with some notable exceptions. For example, birds have lost most KZNFs (Imbeault et al. 2017), such that the median number of ZNF ORFs in birds is a quarter that of other chordates (121 vs. 489); correspondingly, there is no correlation between ZNF and retroelement copy number. Similarly, we observe no correlation across dipteran flies, which possess a large ZNF subfamily known as ZAD-ZNFs; these have been studied in

some depth in *D. melanogaster* and have diverse roles in heterochromatin organization but do not appear to specifically target TEs for silencing (Chung et al. 2007; Kasinathan et al. 2020; Baumgartner et al. 2022). Notwithstanding these exceptions, our analysis indicates that ZNF and TE copy numbers are positively correlated across the breadth of metazoans.

Using our refined gene annotations of FZNFs in cyprinids and improved, TE class-aware estimates of genomic TE content generated with dnaPipeTE (Goubert et al. 2015), we find that the correlation between ZNFs and TEs persists (Fig. 2B,C). Furthermore, we

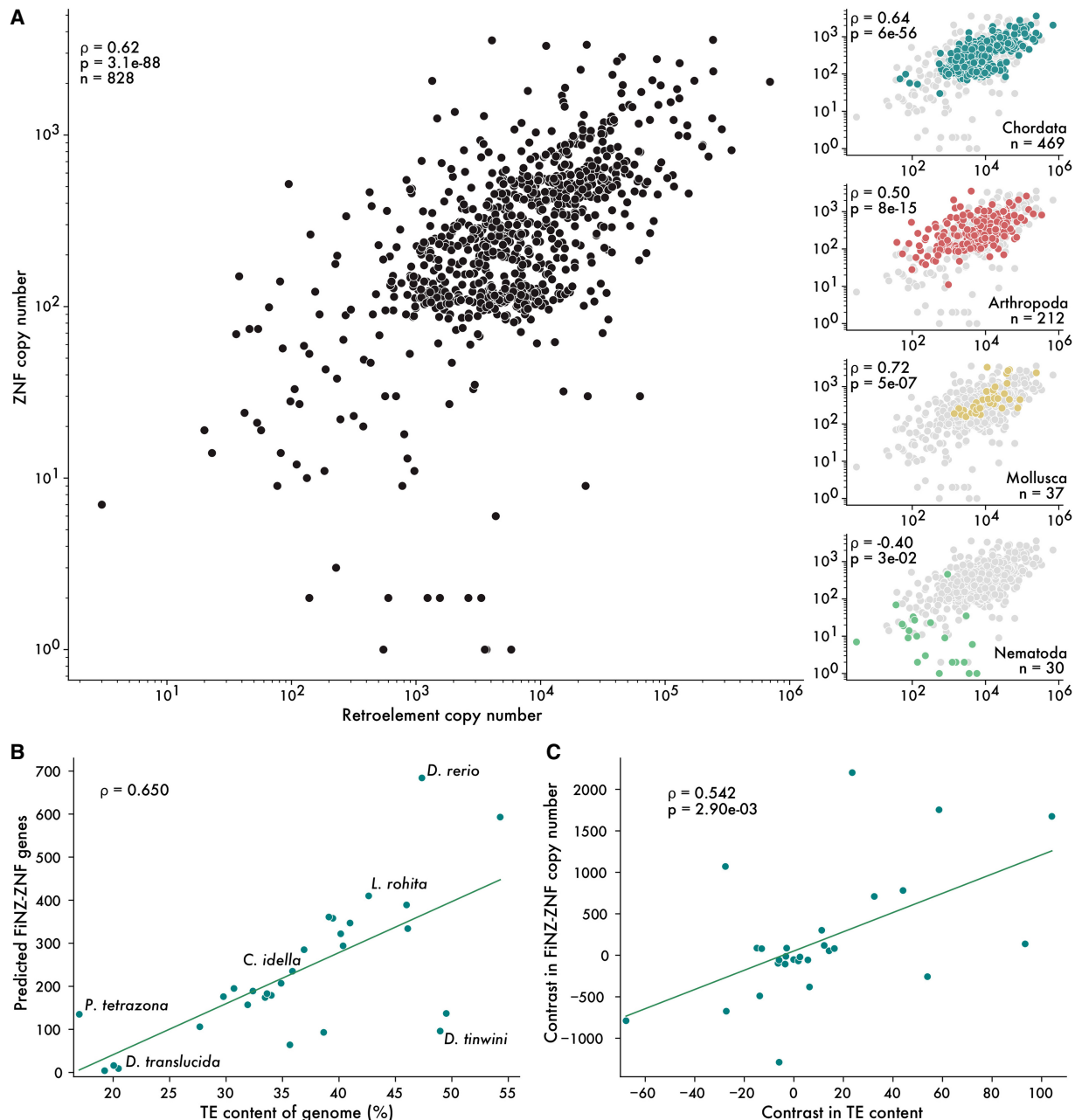


Figure 2. ZNF copy number correlates with genomic TE content. (A) Spearman's rank correlations between the estimated number of retroelement insertions and ZNF ORFs. Each point is a representative species from a taxonomic family. Subplots on the *right* show correlations for individual phyla. (B) Uncorrected Spearman's rank correlation between TE content and FZNF copy number in cyprinid fish. (C) Spearman's rank correlation on phylogenetically independent contrasts.

find that the relationship is not specific to ZNFs and retroelements but is also observed when considering DNA transposon content (Supplemental Fig. 5). These results suggest that the relationship between FZNFs and TEs is not unduly influenced by gene annotation quality and, moreover, applies to a broad range of TEs, not only transcriptionally active retroelements.

We considered two nonmutually exclusive explanations for the correlation between TEs and ZNFs. First, ZNFs may physically interact with TEs, as is the case with KZNFs (Ecco et al. 2017; Imbeault et al. 2017). Second, the TE content of the genome may have an enhancing effect on the rate of gene duplication, such that genomes with large TE content will tend to have larger gene families. The latter model is conceivable because TE accumulation is thought to affect rates of gene turnover via ectopic recombination in a copy number-dependent fashion (Charlesworth and Langley 1989; Jurka et al. 2004; Tollis and Boissinot 2012). One might predict that TE content would positively correlate not only with the copy number of ZNFs but also that of other expansive gene families. We therefore used olfactory receptors as a control case; olfactory receptors are membrane proteins known to turn over rapidly across species but are not expected to interact physically with TEs. In a sample of representative species from 103 mammalian families, we observed no significant correlation between counts of ZNF and olfactory receptor ORFs (Supplemental Fig. 6). Although this analysis does not rule out the possibility that TE content has an impact on the rate of gene duplication, it does imply that the effect is insufficient to explain the correlation observed between ZNFs and TEs.

ZNFs are predicted to bind TE sequences

Hypothesizing that ZNF proteins interact with TEs through sequence-specific recognition of their DNA sequences, we sought to test whether ZNF binding motifs are enriched in TE sequences. Defining the binding specificity of individual ZNF genes experimentally is labor intensive, and the most successful efforts to date have involved large ChIP-based screens in human and mouse cell lines (Imbeault et al. 2017; Wolf et al. 2020). Stable cell lines are not available for most animal species, but computational methods have been developed to predict ZNF binding motifs directly from their protein sequence (Kaplan et al. 2005; Molparia et al. 2010; Persikov and Singh 2014; Najafabadi et al. 2015). Currently, these methods are not sufficient to predict binding specificity with a high degree of accuracy, making the design of reporter assays for ZNF binding impractical. However, we reasoned that if ZNFs preferentially bind TEs, then this preference should be statistically detectable when analyzing large numbers of genes. We therefore used the approach by Najafabadi et al. (2015) to predict DNA-binding motifs for all ZNF ORFs from seven metazoan species with curated libraries of TE consensus sequences: four species with a relatively large repertoire of poorly characterized ZNFs (octopus, zebrafish, rice weevil, and sea urchin), two positive control species known to use ZNFs to target TEs (human and mouse), and an expected negative control species, *D. melanogaster*, whose ZNFs are apparently not targeting TEs directly (Supplemental Data 3).

First, we used previously published ChIP-exo data for human KZNFs (Imbeault et al. 2017) to confirm that computationally predicted DNA-binding motifs were similar to those experimentally determined. Searching for each of 236 experimentally determined KZNF binding sites in the set of predicted binding motifs, we found that 11 (5%) produced significant matches to their predicted coun-

terpart (Q -value < 0.05), compared with zero matches after shuffling the predicted motifs. Similar results have been shown by others working with mice (Wolf et al. 2020). As a second positive control, we compared the predicted binding motifs for human and mouse ZNF ORFs to libraries of their consensus TE sequences and observed a significant enrichment of predicted matches within TEs, compared with matches using shuffled motifs (Fig. 3A). This result is expected because KZNFs are known to target TEs and confirms that predicted binding sites recapitulate, at least in part, experimentally obtained motifs. For example, we were able to recapitulate human ZNF320 binding to LTR14a and ZNF483 binding to L1PA7 (Supplemental Fig. 7). These controls show that although computational prediction of ZNF binding specificity remains challenging, it is sufficient to capture a proportion of biologically relevant binding activity.

We therefore turned our attention to the four species whose ZNFs are less studied (zebrafish, octopus, rice weevil, sea urchin). For each species, we predicted binding motifs for all ZNF ORFs and then examined the frequency of these motifs in corresponding libraries of consensus TE sequences (Fig. 3A). Using a false-discovery threshold of $q = 0.05$, we searched for significant matches between our predicted ZNF binding motifs and TE consensus sequence and then compared them to equivalently sized libraries of nonrepetitive genomic sequence. For all four species, we observed a significant enrichment of matches to the TE sequences compared with nonrepetitive sequences (Fig. 3A). In contrast, we do not observe such enrichment for *D. melanogaster*. Rather, for that species, we found a greater percentage of nonrepetitive sequences matched by predicted ZNF binding motifs (Fig. 3A). This finding is consistent with previous functional studies of *Drosophila* ZAD-ZNFs (Kasinathan et al. 2020) and our observation that the copy number of dipteran ZNFs does not correlate with the number of TE copies (Supplemental Data 1). Notably, we also found that in all species both retroelements and DNA transposons were predicted to be targeted by ZNFs (Supplemental Data 3).

As an additional control, we repeated the analysis using shuffled motifs and found in all species that the predicted motifs bound a higher percentage of TE sequences compared with the shuffled motifs (Fig. 3A). We note, however, that shuffled motifs still recognize a substantial fraction of sequences in most species, likely because ZNFs are often predicted to bind simple sequence repeats or homopolymeric tracts that would be unaffected by shuffling (e.g., Fig. 3B). Finally, to test species specificity, we searched the *D. rerio* TE library using human ZNFs and compared these to zebrafish ZNFs. We found that zebrafish ZNFs were far more likely to target zebrafish TEs than were human ZNFs (odds ratio = 29.5, $P \approx 0.0$, Fisher's exact test). These results suggest that the preferred binding of ZNFs to TE sequences is not unique to mammalian KZNFs but is likely a general property of many metazoan ZNF families. However, direct experimental validation of ZNF binding in non-mammalian organisms will be required before we can be confident that these predictions reflect reality.

FZNFs are evolving under positive selection in zebrafish

TEs evolve rapidly and are under strong selection to escape repression by their hosts; many KZNF proteins are in direct competition with TE families, leading to signatures of arm races between the two (Jacobs et al. 2014; Fernandes et al. 2018; Bruno et al. 2019). One such signature is positive selection (i.e., adaptive evolution) acting on the DNA contacting residues of ZNFs, namely, those involved in TE recognition (Emerson and Thomas 2009). We

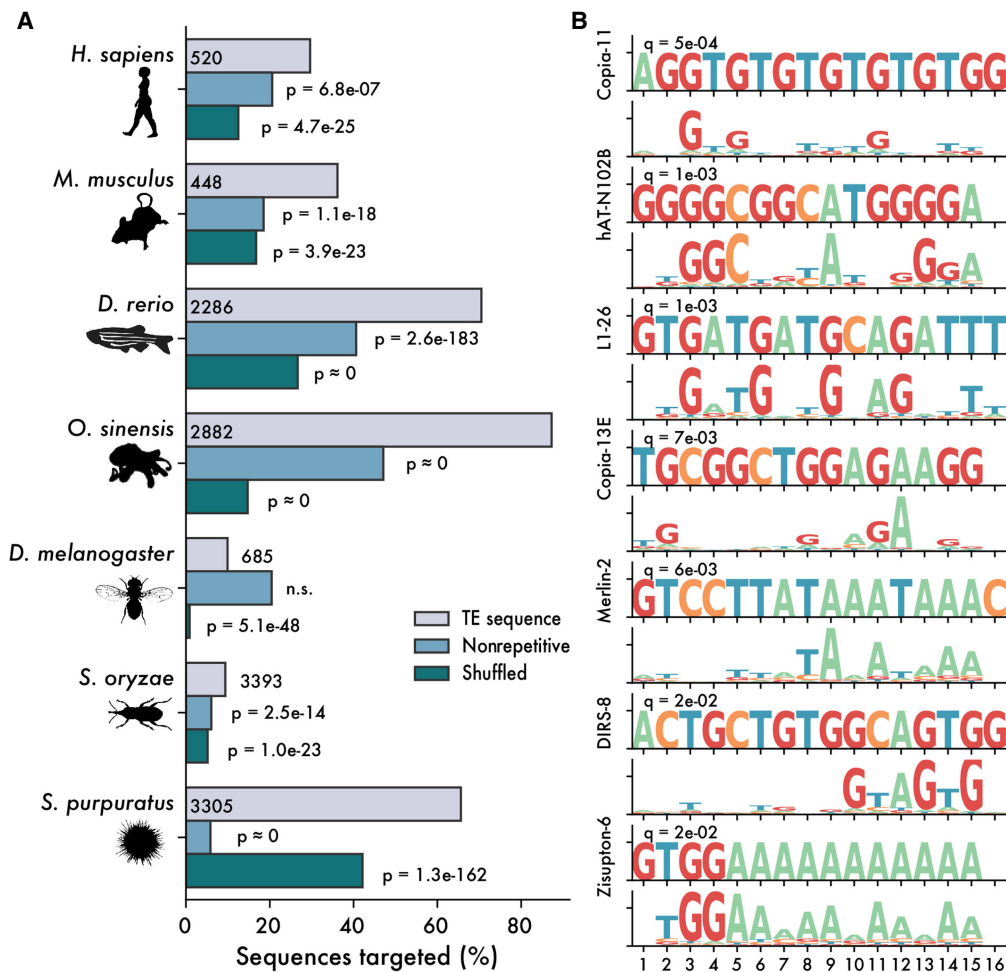


Figure 3. ZNFs are predicted to bind TE sequences. (A) Percentage of sequences matched by predicted ZNF motifs, with a Q-value cutoff of 0.05. Gray bars represent matches to libraries of consensus TE sequences, and blue bars represent matches in equally sized libraries of nonrepetitive genomic sequence. Green bars represent shuffled motifs, searched against consensus TE libraries. P-values are calculated as one-tailed binomial tests for the probability of observing at least x matches in the TE library, using binding frequency to nonrepetitive sequences or with shuffled motifs as a base probability. (B) Selected matches between *Danio rerio* consensus TE sequences and ZNF ORFs.

therefore tested for evidence of positive selection in *D. rerio* FZNFs by comparing rates of synonymous and nonsynonymous substitution (d_N/d_S).

We restricted our analyses to seven clades of recently duplicated paralogs unique to *D. rerio* with a minimum of 10 members (for details, see Methods). Using PAML (Yang 2007), we performed likelihood ratio tests to compare two models of evolution for each clade: one in which the genes were evolving under purifying selection, and one in which some sites were assumed to be under positive selection (M2 vs. M2a). In four of these seven clades, the model featuring positive selection (M2a) was significantly favored (P -value < 0.01 , likelihood ratio tests) (Supplemental Data 4). Furthermore, in two clades we found that base-contacting residues of ZNF domains were significantly enriched for values of $d_N/d_S > 1$ (odds ratio > 2.45 , P -value ≤ 0.05 ; Fisher's exact test). As an alternative to d_N/d_S analyses, which are dependent on accurate gene alignments, we calculated the sequence entropy at positions in the canonical ZNF domain for each species in Figure 3. This revealed that base-contacting residues are among the most variable in sequence, with the possible exception of *D. melanogaster* (Supplemental Fig. 8). These data, combined with the enrichment

of ZNF predicted binding sites in TE sequences, suggest that zebrafish ZNFs, and likely those of many other species, are evolving adaptively to target TEs.

FZNFs are expressed in distinct waves during embryogenesis

TEs—retroelements especially—are highly active during metazoan embryogenesis, as this provides them with the opportunity to transpose in the cells giving rise to the adult germline, thus ensuring the vertical inheritance of new insertions (Haig 2016; Rodriguez-Terrones and Torres-Padilla 2018; Ansaloni et al. 2019; Chang et al. 2022). Previous work has shown that zebrafish ZNFs are expressed at the onset of zygotic genome activation (ZGA), mirroring the pattern of KZNFs in human and mice and consistent with a potential role in TE silencing (White et al. 2017; Pontis et al. 2019; Hadzhiev et al. 2023). We sought to further explore the timing of FZNF expression, predicting that if they are deployed to repress TEs, their expression should overlap with that of TEs. Using our de novo FZNF annotations, we remapped previously published RNA sequencing data covering the first 24 h of zebrafish development (White et al. 2017). Setting a lower limit of 0.5

transcripts per million (TPM) to call genes as expressed, we find that approximately half of all FZNFs (306 out of 684) are expressed during early development.

Recently published work identified two distinct waves of ZNF expression in zebrafish, named “sharp peak” and “broad peak” (Hadzhiev et al. 2019, 2023). Sharp peak FZNFs share a distinct promoter architecture consisting of a clearly identifiable TATA-box and are expressed in the minor wave of ZGA, followed shortly after by TATA-less broad peak FZNFs in the major wave. With our updated FZNF annotations, we recapitulated these findings, observing the expression of 80 sharp peak and 204 broad peak genes, hereafter termed “early” and “late” based on the timing of their expression (Fig. 4A). The timing of these peaks overlaps with that of most TE families and, in the case of early-stage FZNFs, precedes almost all zygotic TE transcription.

We also found a small subset of approximately 22 FZNF genes with maternally deposited transcripts (Fig. 4A). To our knowledge, these have not previously been described, likely owing to their incomplete annotation in Ensembl and RefSeq gene sets. Furthermore, maternally deposited mRNAs in zebrafish and other animals are, initially, only partially polyadenylated (Potireddy et al. 2006; Cui et al. 2013; Winata et al. 2018) and are therefore difficult to detect when using poly(A)-selected RNA-seq platforms. As the White et al. (2017) data falls into this category, we remapped rRNA-depleted reads from a second data set covering early embryogenesis (Winata et al. 2018), which confirmed that these maternally deposited transcripts are indeed present in both the egg and one-cell zygote (Fig. 4B).

Given the markedly different expression pattern of maternally deposited FZNFs, we looked for other features differentiating them from their zygotically expressed counterparts. Although the majority of *D. rerio* FZNFs are recently duplicated, particularly early-stage FZNFs, those that are maternally deposited are conserved across species and are significantly older (Fig. 4C; Supplemental Data 5). Moreover, they are physically colocalized in the subtelomeric region of Chr 4q (Fig. 4D), which lacks the repeat density that characterizes much of Chr 4q (Chang et al. 2022). The age of maternally deposited genes, as well as their degree of species conservation, is hard to reconcile with a role in TE targeting, as TEs turn over rapidly and are frequently species-specific. Rather, the deeper conservation and physical clustering of these maternally deposited FZNFs suggests a role in development that differs from that of early- and late-stage FZNFs.

FZNFs repress LTR retroelement expression during ZGA

Next, we turned to the function of FZNF genes. Based on their embryonic expression, predicted binding specificity, and strong correlation with TE content, we predicted that FZNF genes play a role in TE silencing. If true, then knocking down their expression ought to lead to increased TE expression. To test this, we used a Morpholino oligonucleotide (MO) strategy aimed at blocking translation of most FZNF proteins simultaneously. By aligning the first exon of the FZNF annotations from zebrafish, we were able to design an MO predicted to target the translation start site of 447 out of 684 annotated FZNFs (Fig. 5A). We injected zebrafish embryos with either this MO or a scrambled control and then compared gene and TE expression using RNA-seq. Embryos were collected for sequencing at shield stage, as this corresponds with a period at which many TE families and most FZNFs are robustly expressed (Fig. 4A).

After sequencing and read mapping, we performed quality controls before comparing TE expression between treatment and

control groups. First, we performed principal component analyses of batch-corrected samples, which showed clear separation between the FZNF knockdown and control groups (Supplemental Fig. 9A). Using a false-discovery rate threshold of 0.05 and an absolute \log_2 fold change of at least 0.32 to call differentially expressed genes and TE families, we observed a total of 999 differentially expressed genes, excluding FZNFs and TEs (Supplemental Data 6). Among these, a Gene Ontology enrichment analysis revealed no notable enrichment for biological processes beyond general terms related to embryonic development, metabolic processes, and translation (Supplemental Data 7). Importantly, we did not observe any enrichment for stress response pathways, nor did we observe a significant change in *tp53* levels (\log_2 fold change = -0.01), which can cause off-target effects in MO knockdowns (Bedell et al. 2011).

As the expression of TEs and FZNF genes is highly dynamic during the first 24 h of zebrafish development (Laue et al. 2019; Guo et al. 2021; Chang et al. 2022), we sought to ensure that the embryos were staged appropriately, as differences in the relative stage of FZNF knockdown and control groups could lead to artifactual observations of differential expression. We therefore compared our samples with samples from White et al. (2017), covering 50% epiboly, shield stage, and 75% epiboly. Treating these samples as additional, batch-corrected controls, we found that our samples clustered with shield stage samples from White et al. (2017), indicating that our staging is accurate (Supplemental Fig. 9B,C).

To assess whether FZNFs are involved in TE silencing, we asked whether TE expression was up-regulated in the FZNF knockdown group. Looking at the overall number of differentially expressed TE families, we found that TEs were significantly more likely to be up-regulated than down-regulated, consistent with derepression following FZNF knockdown (odds ratio = 3.78, P -value = 6.1×10^{-8} , Fisher's exact test) (Fig. 5B; Supplemental Fig. 9D,E). Separating by TE class, we found that this trend was driven predominantly by LTR elements (odds ratio = 6.56, P -value = 1.2×10^{-8} , Fisher's exact test) (Supplemental Fig. 10A). Because LTRs include many of the youngest TE families in the zebrafish genome (Chang et al. 2022), we reasoned that their enrichment among up-regulated genes and TE families may reflect a tendency for young, active families to be most intensively targeted by FZNFs. We therefore compared young and old TE families to each other, defining the former as those whose insertions averaged fewer than 0.01 substitutions per site (Chang et al. 2022). This analysis revealed that the magnitude of derepression following FZNF knockdown was significantly greater among young families (P -value = 5.5×10^{-5} , Wilcoxon rank-sum test) and that this trend held across different TE classes (Fig. 5C). Thus, FZNF knockdown leads to significant derepression of young, potentially active TE families.

In addition to age, we also looked at the timing of expression, because by collecting at “shield” stage (~6 h post fertilization), we would not be able to observe the effects on those TEs whose expression peaked at later time points. Using hierarchical clustering to separate TEs into groups with expression peaks either before or after shield stage, we found that those TE families that were expressed earlier were significantly more up-regulated than those expressed later (Supplemental Fig. 10B).

The genes most up-regulated by FZNF knockdown were FZNFs themselves, for example, *g204* and *g422* (Fig. 5D). Although we cannot currently rule out the possibility that MOs inhibit the degradation of their target sequences, to our knowledge, such effects have not previously been reported in the literature. Alternatively, these results suggest that FZNF genes are themselves targeted for

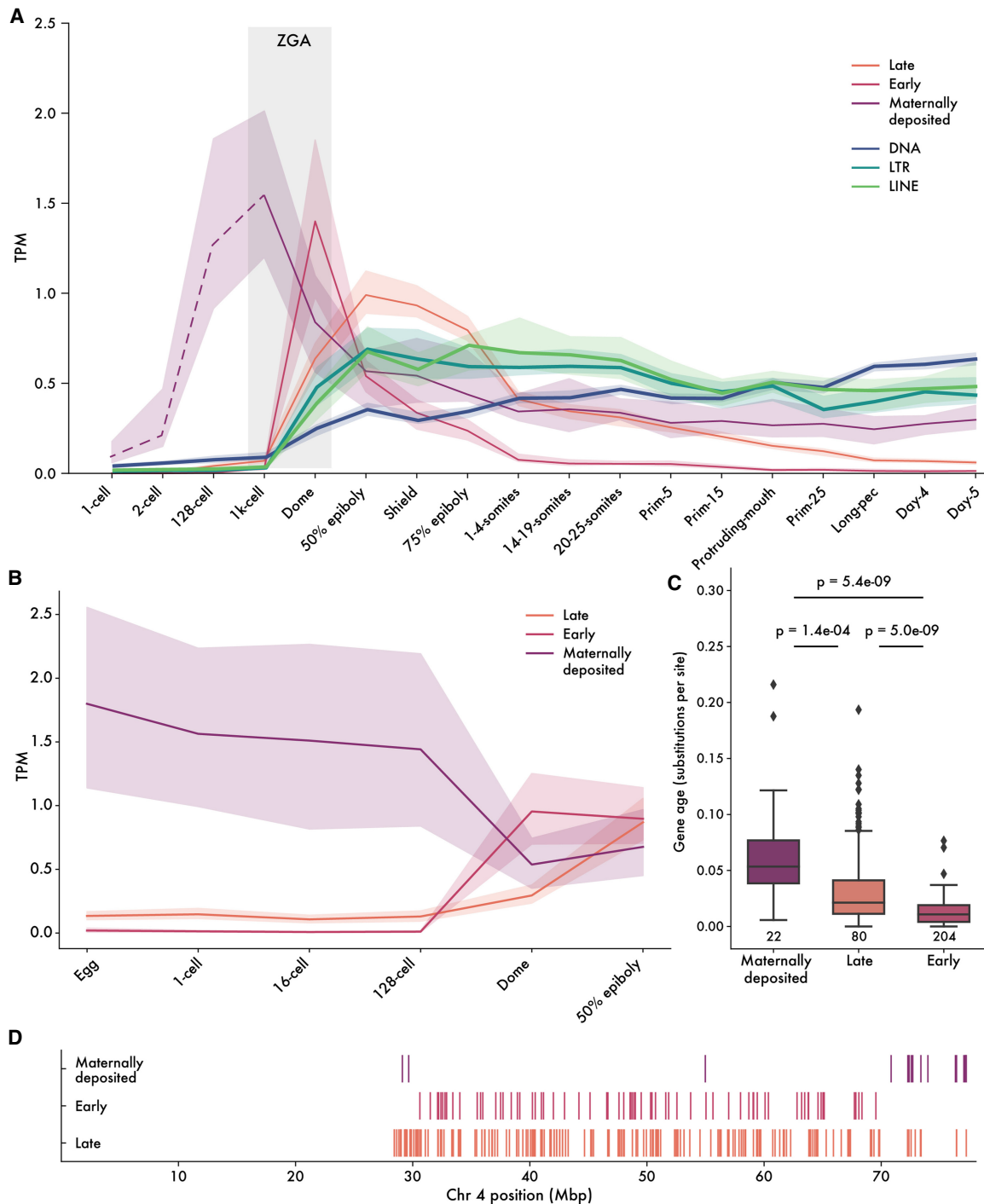


Figure 4. Embryonic expression of FZNF genes in zebrafish. (A) Median expression trajectory of 306 genes with expression of at least 0.5 TPM in at least one stage, separated by cluster. Shaded borders represent 90% confidence intervals. Light gray bar shows approximate timing of ZGA. Expression trajectory of maternally deposited FZNFs before the 1000-cell stage is dashed to reflect the fact that poly(A)-selected RNA sequencing is not an accurate reflection of true mRNA levels before ZGA. (B) FZNF expression from rRNA-depleted reads, before ZGA. (C) Maternally deposited genes are significantly older, whereas early-stage FZNFs are young and generally unique to *D. rerio*. Ages were quantified as branch lengths on a neighbor-joining phylogenetic tree generated from pairwise distances. (D) The majority of maternally deposited FZNF genes are located in a specific, subtelomeric cluster. Each vertical line represents the midpoint of an expressed FZNF gene.

repression by FZNF proteins, thus forming a negative feedback loop. In this way, their expression would be robustly switched off once sufficient FZNF protein had been generated, without the

need for additional regulatory control. Such negative feedback loops have recently been documented for mammalian KZNFs (Pontis et al. 2022).

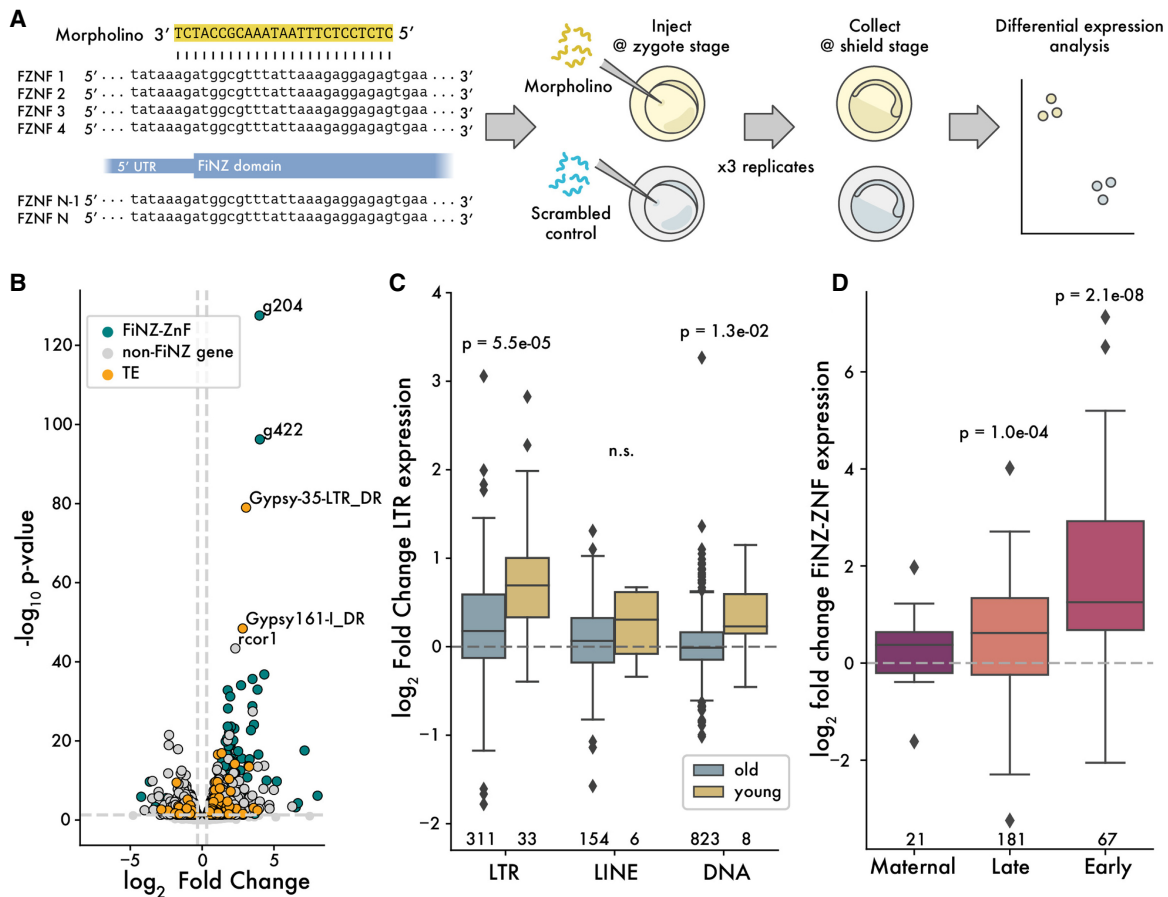


Figure 5. Up-regulation of LTR retroelements in response to FZNF knockdown. (A) Cartoon showing design of MO-based FZNF knockdown experiment. (B) \log_2 fold change versus P -value, comparing FiNZ translation blocking MO to scrambled MO control. (C) \log_2 fold change comparison between old and young TE families, with “young” being defined as those whose median insertion age was less than 0.01 substitutions per site. P -values are Wilcoxon rank-sum tests comparing old and young in each TE class. (D) \log_2 fold change of different FZNF categories. P -values are binomial tests for each category, assessing likelihood of seeing at least x up-regulated genes, assuming a null hypothesis in which there is a 50% chance of being either up- or down-regulated. Numbers below boxes in C and D give the number of data points per box.

Discussion

In this work, we observed a deep coevolutionary relationship between TEs and ZNFs that spans the breadth of metazoans. This is most clearly seen in the strong positive correlation between TE and ZNF copy numbers, which persists across at least 750 million years of animal evolution and is independent of the accessory domains that characterize different tandem repeat ZNF families. We also found that ZNFs from a range of metazoans are predicted to bind TE sequences and, in the case of FZNFs in zebrafish, are evolving under positive selection acting on the DNA-binding ZNF domains. Finally, through simultaneous knockdown of a large subset of FZNFs during zebrafish embryogenesis, we showed that this family represses young, recently active TEs.

The extensively duplicated ZNF families of metazoans are rarely observed in other eukaryotic taxa, where instead, small-RNAs acting in concert with Argonaute proteins are the primary drivers of TE heterochromatinization (Hall et al. 2002; Volpe et al. 2002; Najafabadi et al. 2017; Allshire and Madhani 2018; Heger et al. 2020). Thus, ZNF-based TE targeting is likely to be a metazoan innovation. However, there are several metazoan groups whose ZNF families appear to behave atypically. For example, we found no positive correlations between ZNFs and retroelements

in dipteran flies or nematode worms, nor did we observe a binding preference for TEs in *Drosophila* ZNFs. However, in *Drosophila* the piRNA system (itself a branch of the Argonaute-siRNA system) is directly involved in pretranscriptional silencing of TEs during embryogenesis, with *piwi* knockdowns resulting in a failure of H3K9me3 to establish over otherwise silenced TE loci (Sienski et al. 2012; Fabry et al. 2021; Wei et al. 2021). Similarly, in nematode worms, which typically have few tandem-repeat C2H2 ZNF genes, Argonaute genes have been extensively duplicated (Buck and Blaxter 2013). Although research into the function of these Argonaute homologs remains in the early stages (Seroussi et al. 2023), both the dipteran and nematode cases point to potential functional overlap between the Argonaute/piRNA system and ZNFs.

It is still an open question whether silencing of TE transcription per se, is the primary function of metazoan ZNFs. In addition to their TE silencing activities, KZNFs frequently target non-TE sequences, such as satellite repeats, gene promoters, and even other KZNFs (Imbeault et al. 2017). There are also examples in both mouse and human of conserved clusters of KZNFs that target TEs younger than themselves, implying they had targets predating the emergence of their current TE families (Wolf et al. 2015; 2020; Imbeault et al. 2017). These observations have led others

to propose that, rather than simply repressing TEs, KZNFs play an important role in their domestication by taming their *cis*-regulatory activities and facilitating their integration into existing gene regulatory networks (Ecco et al. 2017; Bruno et al. 2019; Pontis et al. 2019). Under this model, KZNFs act as tolerogenic agents promoting the evolution of species-specific gene regulatory networks. Although likely true, this feature cannot be the initial driver of their rapid evolution, because TE domestication is a process that takes place over many generations, whereas selection primarily acts on currently extant individuals (Sniegowski and Murphy 2006).

Although depletion of both KZNFs and FZNFs unleash TE expression, there are several paradoxical observations that challenge the idea of ZNF as an adaptive defense system against TE invasion. First, ZNFs evolve slowly relative to the speed at which novel TE families can establish and propagate within a species. For example, the DNA transposon P-element spread throughout all wild populations of *D. melanogaster* in the space of ~30 yr, and the retrovirus KoRV has been spreading in koalas exogenously and endogenously over a similar period (Anxolabéhère et al. 1985; Tarlinton et al. 2006). In contrast, observations in primates imply that the emergence of KZNF repressors takes place over millions of years (Ecco et al. 2017; Imbeault et al. 2017). In contrast, the piRNA system is able to adapt almost immediately to new TEs, as TE insertions themselves form the raw material for piRNAs—indeed, in the case of both KoRV and P-element, piRNAs have already emerged to mitigate their deleterious effects (Brennecke et al. 2008; Yu et al. 2019). Thus, with respect to their ability to quickly respond to novel TEs, the piRNA system appears better suited as a front-line defense than ZNFs.

We therefore propose an alternative to transcriptional repressor hypothesis, in which metazoan ZNFs instead serve as genome stabilizers, protecting DNA from ectopic recombination by nucleating heterochromatin to repetitive regions of the genome. Although heterochromatin is primarily associated with gene silencing, it also has a strong suppressive effect on homologous recombination and is essential for maintaining genome stability (for reviews, see Grewal 2010; Sasaki et al. 2010; Kent et al. 2017; Allshire and Madhani 2018). TE-mediated chromosomal rearrangements are highly deleterious and associated with several hereditary diseases (Callinan and Batzer 2006), both in germline and somatic tissues, as evidenced by the large number of human cancers associated with TE-induced structural variants (Anwar et al. 2017). Accordingly, the propensity of TEs to cause deleterious chromosomal rearrangements is thought to be a critical factor preventing their accumulation in genomes, particularly because the theoretical probability of recombination between TE insertions increases quadratically with copy number (Charlesworth and Charlesworth 1983; Langley et al. 1988). This hypothesis—that selection against ectopic recombination is critical to preventing TE accumulation—has been supported by observations from both empirical and simulated data (Petrov 2003; Song and Boissinot 2007; Dolgin and Charlesworth 2008; Kofler et al. 2012), although others have drawn different conclusions (Duret et al. 2000; Wright et al. 2003).

Thus, the use of sequence-specific ZNFs to deposit heterochromatin over discrete TE families might explain in part why some organisms can sustain large-scale amplification of TE families, as seen in octopuses, weevils, sea urchins, cyprinid fish, and mammals, to name a few highlighted in this study. Moreover, both KZNFs and FZNFs, along with other highly duplicated gene families, are also found in regions of the genome marked by H3K9me3, further suggesting that stabilization of repeats, rather

than suppression of TE transcription, is the purpose of heterochromatin deposition at these locations (Hawkins et al. 2010; Kundaje et al. 2015). If genome stabilization is the principal function of most metazoan ZNFs, then it relieves the requirement for rapid adaptation to novel TEs, because the effect of TEs on rates of nonallelic homologous recombination are not felt until the TEs in question have reached significant copy number. Similarly, it would explain why satellite sequences or nonautonomous DNA elements, which do not need to be transcribed to amplify, appear to be targets of many ZNFs.

Moving forward, the FZNF family will be an important model for understanding the function and evolution of metazoan ZNFs, and many questions raised in this study remain unanswered. For example, what is the mechanism by which FZNFs exert their repressive effect? Although zebrafish FZNFs are, in aggregate, predicted to target TE sequences (Fig. 3), we have not yet validated the binding of individual ZNF-TE pairs, and it therefore remains possible that they achieve silencing via a mechanism other than DNA-binding. Similarly, we do not yet know whether the FiNZ domain itself is capable of recruiting the protein complexes required for heterochromatin formation. Although KRAB and FiNZ have similar sequence composition and secondary structure, the KRAB domain recruits the H3K9me3 writer SETDB1 through an intermediary cofactor, TRIM28, which is absent in zebrafish. However, two close relatives of TRIM28, namely, TRIM24 and TRIM33, are present in zebrafish and have been shown to silence endogenous retroviruses in mice (Boudinot et al. 2011; Margalit et al. 2020). Furthermore, structural predictions suggest that TRIM33 physically interacts with distantly related “KRAB-like” ZNF sequences present in fish (Carotti et al. 2022), suggesting that it may be the cofactor of FiNZ.

Much remains to be discovered about the role of ZNF genes in metazoan evolution, but with the wealth of high-quality genome assemblies and improved tools for manipulating gene expression in a variety of organisms, it is now possible to answer many of the long-standing questions about this extraordinary gene family. In this work, we have shown that the coevolutionary relationship between TEs and ZNFs extends well beyond KZNFs and is likely a metazoan innovation for TE silencing and/or stabilization. In zebrafish, FZNFs have independently expanded to repress young, recently active TE families during early embryonic development, and across animals, there are dozens of other such uncharacterized ZNF families. Many of these families are likely to reveal exciting new biology and may have potential uses in the development of genome engineering tools.

Methods

Predicting ZNF ORF copy number

Initial estimates of ZNF ORF copy number were achieved by extracting all possible ORFs from 3221 representative metazoan genomes available from the NCBI Assembly database (<https://www.ncbi.nlm.nih.gov/assembly>), as of December 6, 2021. To facilitate this process, we set minimum and maximum ORF sizes of 375 and 10,000 bp, respectively. The resulting amino acid sequences were searched for potential C2H2 ZNF domains using the *hmmsearch* command line tool from the HMMER suite (Eddy 2011) and the Pfam “zf_C2H2” (Pfam: PF00096) HMM as a query profile (Mistry et al. 2021). Resulting hits were merged if they were within 100 residues of each other and were filtered to exclude sequences with fewer than five ZNF domains in total.

Annotation of FZNF genes

For improved annotations of FZNF genes in 32 cypriniform fish and a milkfish outgroup, we first identified genomic regions containing candidate FZNF genes using BLAST to search for matches with a consensus FinZ domain sequence. This consensus sequence was generated from a set of *D. rerio* FZNF genes previously identified as being expressed during development (White et al. 2017). Using the “PPX” module of AUGUSTUS v3.3 (Keller et al. 2011; Hoff and Stanke 2018), we performed a first pass of the genomic regions identified in our set of cypriniform species, using a protein profile generated from the previously mentioned set of *D. rerio* annotations. To avoid biasing our search toward zebrafish, we used the results of this round of annotation to generate a new FZNF profile generated by sampling genes from all species. We repeated the above procedure once and performed a final filtering step with HMMER (Eddy 2011) to identify annotated genes containing both the FinZ domain and ZNF domain (PFAM: PF00096).

For *D. rerio* specifically, we produced a separate, high-quality set of gene predictions by retraining AUGUSTUS specifically for FZNFs using a manually curated set of Ensembl predictions. This allowed us to generate predictions for 5' and 3' untranslated regions, for use downstream in analyses of gene expression. We explored the effect of including transcripts generated by Trinity (Grabherr et al. 2011) as hints for AUGUSTUS but found that these reduced the quality of resulting annotations, likely as a result of the difficulty of assembling accurate transcripts for such repetitive genes. Full parameter details are available online.

Predicting genomic TE content

We used two approaches to estimate genomic TE content. In the first, we focused on retroelements, using an identical approach to that which we used to estimate ZNF ORF number. In place of the Pfam zf_C2H2 domain, we instead used a selection of reverse transcriptase and RNase H profiles (PF00078, PF07727, PF13456, PF00075, PF17917, PF17919). Resulting hits were merged as performed for ZNFs, but no restrictions were placed on the number of hits per ORF. For more careful estimates of TE content for use when focusing on FZNF genes, we estimated the TE-derived proportion of genomes using dnaPipeTE (Goubert et al. 2015), a tool that offers considerable speed improvements over RepeatMasker/Modeller with comparable accuracy. This program requires short reads and genome size as input: For the former, we simulated reads using ART (Huang et al. 2012), and for the latter, we used genome assembly size as a proxy for true genome size.

Cypriniformes phylogeny

To generate a phylogenetic tree of the Cypriniformes order (required to account for phylogenetic nonindependence in comparative genomic analyses), we first used the Actinopterygii-specific BUSCO database to extract intact single-copy orthologues from our set of 33 species. We then selected those protein sequences found in at least 10 out of those species, producing a final set of 3581 proteins. These were aligned separately using MAFFT v7.475 with the parameters `-globalpair` and `-maxiterate 1000` (i.e., ginsi) (Katoh and Standley 2013) and trimmed to remove large insertions or poorly aligned regions using trimAl v1.4.rev15 with the `-automated1` parameter set (Capella-Gutiérrez et al. 2009). All resulting files were combined to produce a concatenated supergene alignment, which was used as input when generating a time-calibrated phylogeny.

IQ-TREE v2.0.6 was used to generate a maximum likelihood tree from a partitioned analysis of the supergene alignment (Chernomor et al. 2016; Hoang et al. 2018; Minh et al. 2020).

Ultrafast bootstraps and Shimodaira–Hasegawa approximate likelihood ratio tests were used to determine branch support values, with 5000 and 1000 replicates, respectively. A time-calibrated tree was generated using the integrated LSD2 module (To et al. 2016), constraining the date of the split between Gonorynchiformes and Cypriniformes (i.e., between *Chanos chanos* and all other species) to 162 million years ago, based on a recent comprehensive phylogeny of teleost fish (Hughes et al. 2018).

Predicting ZNF binding sites

To predict ZNF binding sites, we made use of the previously published tool, ZiFRC, with ZNF ORFs identified from our initial genome searches as input (Najafabadi et al. 2015; 2017). To test for enrichment of predicted ZNF binding sites in various sequence sets, we used FIMO and Tomtom from the MEME suite of motif enrichment tools (v5.3.3) (Bailey et al. 2015). First, Tomtom was used to compare our predicted motifs to motifs obtained from human ChIP-exo data sets (Imbeault et al. 2017). Following this positive control, we used FIMO to calculate motif enrichment in three separate cases for each species.

In the first case, we searched our library of predicted ZNF binding sites against libraries of TE consensus sequences. For human, mouse, fruit fly, and zebrafish, these libraries were taken directly from Repbase v26.10; in the case of sea urchin and rice weevil, they were provided by direct communication with the investigators of previously published studies on the TE content of these species (Panyushev et al. 2021; Parisot et al. 2021); and, in the case of octopus, we generated a TE consensus sequence library using RepeatModeler v2.0.1 (Flynn et al. 2019).

For the second case, we searched a library of randomly selected nonrepetitive genomic sequences, selected using BEDTools v2.30.0 (Quinlan and Hall 2010); this was performed in a manner ensuring that, for each species, the resulting library size and distribution of sequence lengths were identical to those of the TE sequence library. In the third case, we searched the TE consensus sequences again, this time using shuffled binding motifs. Finally, we performed a cross-species comparison between zebrafish and human using FIMO.

Testing for positive selection

To test for evidence of positive selection in FZNF genes, we first searched for recently duplicated in-paralogs. As multiple sequence alignments of ZNF genes can be unreliable, we calculated pairwise Needleman–Wunsch alignments between all pairs of 6732 annotated genes from 29 cypriniform species with at least one FZNF gene, using the “needle” command line tool from the EMBOSS suite, v6.6.0 (Rice et al. 2000). The resulting pairwise distance matrix was used as input to generate a neighbor-joining phylogenetic tree, using the EMBOSS “fneighbor” tool. This gene tree was used to identify clades of recently duplicated in-paralogs within the cypriniform species tree. We selected seven clades from *D. rerio*, in which each clade contained at least 10 members, and all members were within 30 amino acids of each other in length. Each one of these clades was realigned with Prank, using a codon substitution matrix (Kosiol et al. 2007; Löytynoja 2014), and all gaps in the resulting alignment were removed using trimAl v1.4.rev15 (Capella-Gutiérrez et al. 2009). Finally, guide trees were generated for each clade using IQ-TREE v2.0.6 with a codon substitution model (Minh et al. 2020).

To test for evidence of positive selection, we used the “codeml” tool from the PAML4 software suite (Yang 2007). We compared two sets of nested site models: first, M1a (NearlyNeutral) and M2a (PositiveSelection), and second, M7

(beta) and M8 (beta & ω). The transition-to-transversion ratio, κ , was calculated from the guide tree, whereas the d_N/d_S rate ratio, ω (i.e., the parameter being tested), was estimated from the data. Likelihood ratio tests with two degrees of freedom were used to compare M1a to M2a, and M7 to M8, respectively.

Reanalysis of zebrafish developmental transcriptome data

To assess FZNF gene and TE expression, we first used STAR (v2.7.5a) (Dobin et al. 2013) in multimapping mode to realign reads from White et al. (2017) to the *D. rerio* GRCz11 reference genome, setting the maximum intron/read pair gap size at 500,000. To count reads from genes and TEs simultaneously, we used TEtranscripts-2.2.1's TEcount tool in multimapper mode, with a GTF file combining Ensembl gene annotations with our FZNF annotations, along with TE locations as provided by Chang et al. (2022). We used GTFtools v0.8.0 (Li et al. 2022) to calculate the median length of each gene and used these measurements to calculate the TPM for each gene/TE, allowing comparison of expression across stages. Specifically, for each transcript we calculate the sum of the union of exons. For genes, we then take the median length of transcripts associated with each gene, whereas for TEs we use the sum of all insertions in the genome, thus partially controlling for insertion copy number.

Morpholino knockdown of FZNF translation

We ordered a translation blocking MO with the sequence "CTCTCCTCTTAATAAACGCCATCT" from Gene Tools, which was sufficient to target 447 of our predicted FZNF mRNA transcripts with no mismatches. Wild-type, Tübingen strain zebrafish were maintained on a 14-/10-h light–dark schedule at 28°C. Individual breeding pairs of male and female fish were separated overnight to induce naturally synchronized spawning and fertilization the following morning. To knockdown FZNF expression, we injected 4.5 ng of the custom MO into fertilized *D. rerio* zygotes, and in parallel, we injected embryos with the same quantity of a randomized control, consisting of a mixture of up to 4²⁵ possible sequences. Embryos were collected at shield stage and stored at –20°C immediately after collection. Each pair of injections was repeated in triplicate, such that 23 treatment and 19 control embryos were in the first batch, 30 treatment and 30 control in the second, and 31 and 30, respectively, in the third. Total RNA extraction was performed using the Qiagen RNeasy kit.

RNA sequencing

We outsourced cDNA library preparation and sequencing to Novogene. Samples were shipped in dry ice, and after quality controls, strand-specific libraries were prepared using poly(A) enrichment. These libraries were pooled and sequenced in 150-bp paired-end mode using the NovaSeq 6000 platform. After receiving sequencing data, we performed quality assessment using FastQC (v0.11.9) (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and then mapped and counted reads according to the protocol previously described for reanalysis of White et al. (2017) data. Following this, differential expression analyses were performed using the DESeq2 R package v1.26.0 (Love et al. 2014).

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO); <https://www.ncbi.nlm.nih.gov/geo/> under accession number GSE229157. Scripts used in the generation and analysis of data for this project can be found at GitHub (<https://github.com/>

[jonathan-wells/metazoan-znfs](https://github.com/jonathan-wells/finz-znf); <https://github.com/jonathan-wells/finz-znf>) and as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank Dr. Joseph Fetcho for zebrafish husbandry support. We thank members of the Feschotte laboratory for valuable feedback and discussion. This work was supported by grant R35-GM122550 from the National Institutes of Health to C.F. J.N.W. is supported by a Human Frontier Science Program long-term fellowship (LT000017/2019-L). N.-C.C. is supported by a Distinguished Scholar Award from the Cornell Center for Vertebrate Genomics. C.C. was supported by a stipend from the Cornell Research Experience for Undergraduates program.

Author contributions: J.N.W. performed all computational analyses. N.-C.C. performed FZNF MO injections. J.M. performed early studies of the FiNZ domain and provided general assistance with “wet-laboratory” experiments. C.C. generated and analyzed preliminary data from surveys of metazoan ZNF and TEs. N.R. performed early annotation of *Danio* genus TEs and performed pilot wet-laboratory experiments. B.J. generated and analyzed preliminary data for testing correlation between TEs and other gene families. J.N.W. designed all experiments with input from N.-C.C., J.M., and C.F. J.N.W. wrote the manuscript and created figures with input from C.F. Initial project conception was by J.N.W., J.M., and C.F.

References

- Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, Brenner S, Ragsdale CW, Rokhsar DS. 2015. The octopus genome and the evolution of cephalopod neural and morphological novelties. *Nature* **524**: 220–224. doi:10.1038/nature14668
- Allshire RC, Madhani HD. 2018. Ten principles of heterochromatin formation and function. *Nat Rev Mol Cell Biol* **19**: 229–244. doi:10.1038/nrm.2017.119
- Altschul SE, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2
- Ansaloni E, Scarpato M, Di Schiavi E, Gustincich S, Sanges R. 2019. Exploratory analysis of transposable elements expression in the *C. elegans* early embryo. *BMC Bioinformatics* **20**: 484. doi:10.1186/s12859-019-3088-7
- Anwar SL, Wulaningsih W, Lehmann U. 2017. Transposable elements in human cancer: causes and consequences of deregulation. *Int J Mol Sci* **18**: 974. doi:10.3390/ijms18050974
- Anxolabéhère D, Nouaud D, Périquet G, Tchen P. 1985. P-element distribution in Eurasian populations of *Drosophila melanogaster*: a genetic and molecular analysis. *Proc Natl Acad Sci* **82**: 5418–5422. doi:10.1073/pnas.82.16.5418
- Aravin AA, Hannon GJ, Brennecke J. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* **318**: 761–764. doi:10.1126/science.1146484
- Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. *Nucleic Acids Res* **43**: W39–W49. doi:10.1093/nar/gkv416
- Baumgartner L, Handler D, Platzer SW, Yu C, Duchek P, Brennecke J. 2022. The *Drosophila* ZAD zinc finger protein Kuperfel guides rhino to piRNA clusters. *eLife* **11**: e80067. doi:10.7554/eLife.80067
- Bedell VM, Westcot SE, Ekker SC. 2011. Lessons from morpholino-based screening in zebrafish. *Brief Funct Genomics* **10**: 181–188. doi:10.1093/bfgp/blr021
- Bellefroid EJ, Poncelet DA, Lecocq PJ, Revelant O, Martial JA. 1991. The evolutionarily conserved Krüppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc Natl Acad Sci* **88**: 3608–3612. doi:10.1073/pnas.88.9.3608
- Boudinot P, van der Aa LM, Jouneau L, Du Pasquier L, Pontarotti P, Briolat V, Benmansour A, Levraud JP. 2011. Origin and evolution of TRIM

- proteins: new insights from the complete TRIM repertoire of zebrafish and pufferfish. *PLoS One* **6**: e22022. doi:10.1371/journal.pone.0022022
- Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* **322**: 1387–1392. doi:10.1126/science.1165171
- Bruno M, Mahgoub M, Macfarlan TS. 2019. The arms race between KRAB-zinc finger proteins and endogenous retroelements and its impact on mammals. *Annu Rev Genet* **53**: 416. doi:10.1146/annurev-genet-112618-043717
- Buck AH, Blaxter M. 2013. Functional diversification of argonautes in nematodes: an expanding universe. *Biochem Soc Trans* **41**: 881–886. doi:10.1042/BST20130086
- Callinan PA, Batzer MA. 2006. Retrotransposable elements and human disease. *Genome Dyn* **1**: 104–115. doi:10.1159/000092503
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973. doi:10.1093/bioinformatics/btp348
- Carotti E, Carducci F, Greco S, Gerdol M, Di Marino D, Perta N, La Teana A, Canapa A, Barucca M, Biscotti MA. 2022. Transcriptional contribution of transposable elements in relation to salinity conditions in teleosts and silencing mechanisms involved. *Int J Mol Sci* **23**: 5215. doi:10.3390/ijms23095215
- Chang NC, Rovira Q, Wells J, Feschotte C, Vaquerizas JM. 2022. Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. *Genome Res* **32**: 1408–1423. doi:10.1101/gr.275655.121
- Charlesworth B, Charlesworth D. 1983. The population dynamics of transposable elements. *Genet Res* **42**: 1–27. doi:10.1017/S0016672300021455
- Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. *Annu Rev Genet* **23**: 251–287. doi:10.1146/annurev.ge.23.120189.001343
- Chernomor O, von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol* **65**: 997–1008. doi:10.1093/sysbio/syw037
- Chung HR, Löhr U, Jäckle H. 2007. Lineage-specific expansion of the zinc finger associated domain ZAD. *Mol Biol Evol* **24**: 1934–1943. doi:10.1093/molbev/msm121
- Cui J, Sartain CV, Pleiss JA, Wolfner MF. 2013. Cytoplasmic polyadenylation is a major mRNA regulator during oogenesis and egg activation in *Drosophila*. *Dev Biol* **383**: 121–131. doi:10.1016/j.ydbio.2013.08.013
- Czech B, Hannon GJ. 2016. One loop to rule them all: the ping-pong cycle and piRNA-guided silencing. *Trends Biochem Sci* **41**: 324–337. doi:10.1016/j.tibs.2015.12.008
- Degnan BM, Vervoort M, Larroux C, Richards GS. 2009. Early evolution of metazoan transcription factors. *Curr Opin Genet Dev* **19**: 591–599. doi:10.1016/j.gde.2009.09.008
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. doi:10.1093/bioinformatics/bts635
- Dolgin ES, Charlesworth B. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**: 2169–2177. doi:10.1534/genetics.107.082743
- Duret L, Marais G, Biémont C. 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. *Genetics* **156**: 1661–1669. doi:10.1093/genetics/156.4.1661
- Ecco G, Imbeault M, Trono D. 2017. KRAB zinc finger proteins. *Development* **144**: 2719–2729. doi:10.1242/dev.132605
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* **7**: e1002195. doi:10.1371/journal.pcbi.1002195
- Emerson RO, Thomas JH. 2009. Adaptive evolution in zinc finger transcription factors. *PLoS Genet* **5**: e1000325. doi:10.1371/journal.pgen.1000325
- Fabry MH, Falconio FA, Joud F, Lythgoe EK, Czech B, Hannon GJ. 2021. Maternally inherited piRNAs direct transient heterochromatin formation at active transposons during early *Drosophila* embryogenesis. *eLife* **10**: e68573. doi:10.7554/eLife.68573
- Fedotova AA, Bonchuk AN, Mogila VA, Georgiev PG. 2017. C2H2 zinc finger proteins: the largest but poorly explored family of higher eukaryotic transcription factors. *Acta Naturae* **9**: 47–58.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat* **125**: 1–15. doi:10.1086/284325
- Fernandes JD, Haussler M, Armstrong J, Tigyi K, Gu J, Filippi N, Pierce J, Thisner T, Angulo P, Katzman S, et al. 2018. KRAB Zinc Finger Proteins coordinate across evolutionary time scales to battle retroelements. bioRxiv doi:10.1101/429563
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2019. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci* **117**: 9451–9457. doi:10.1073/pnas.1921046117
- Goodstadt L, Heger A, Webber C, Ponting CP. 2007. An analysis of the gene complement of a marsupial, *Monodelphis domestica*: evolution of lineage-specific genes and giant chromosomes. *Genome Res* **17**: 969–981. doi:10.1101/gr.6093907
- Goubert C, Modolo L, Vieira C, ValienteMoro C, Mavingui P, Boulesteix M. 2015. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biol Evol* **7**: 1192–1205. doi:10.1093/gbe/evv050
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* **29**: 644–652. doi:10.1038/nbt.1883
- Grewal SI. 2010. RNAi-dependent formation of heterochromatin and its diverse functions. *Curr Opin Genet Dev* **20**: 134–141. doi:10.1016/j.gde.2010.02.003
- Guo Y, Gert KR, Lebedeva S, Potok ME, Wike CL, Grow EJ, Ketting RF, Pauli A, Cairns BR. 2021. Maternally-inherited anti-sense piRNAs antagonize transposon expression in teleost embryos. bioRxiv doi:10.1101/2021.11.03.467172
- Hadzhiev Y, Qureshi HK, Wheatley L, Cooper L, Jasiulewicz A, Van Nguyen H, Wragg JW, Poovathumkadavil D, Conic S, Bajan S, et al. 2019. A cell cycle-coordinated polymerase II transcription compartment encompasses gene expression before global genome activation. *Nat Commun* **10**: 691. doi:10.1038/s41467-019-08487-5
- Hadzhiev Y, Wheatley L, Cooper L, Ansaloni F, Whalley C, Chen Z, Finaurini S, Gustincich S, Sanges R, Burgess S, et al. 2023. The miR-430 locus with extreme promoter density forms a transcription body during the minor wave of zygotic genome activation. *Dev Cell* **58**: 155–170.e8. doi:10.1016/j.devcel.2022.12.007
- Haig D. 2016. Transposable elements: self-seekers of the germline, team-players of the soma. *Bioessays* **38**: 1158–1166. doi:10.1002/bies.201600125
- Hall IM, Shankaranarayana GD, Noma K-I, Ayoub N, Cohen A, Grewal SIS. 2002. Establishment and maintenance of a heterochromatin domain. *Science* **297**: 2232–2237. doi:10.1126/science.1076466
- Haring NL, van Bree EJ, Jordaan WS, Roels JRE, Sotomayor GC, Hey TM, White FTG, Galland MD, Smidt MP, Jacobs FMJ. 2021. ZNF91 deletion in human embryonic stem cells leads to ectopic activation of SVA retrotransposons and up-regulation of KRAB zinc finger gene clusters. *Genome Res* **31**: 551–563. doi:10.1101/gr.265348.120
- Hawkins RD, Hon GC, Lee LK, Ngo QM, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, et al. 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**: 479–491. doi:10.1016/j.stem.2010.03.018
- Heger P, Zheng W, Rottmann A, Panfilio KA, Wiehe T. 2020. The genetic factors of bilaterian evolution. *eLife* **9**: e45530. doi:10.7554/eLife.45530
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* **35**: 518–522. doi:10.1093/molbev/msx281
- Hoff KJ, Stanke M. 2018. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc Bioinformatics* **65**: e57. doi:10.1002/cpbi.57
- Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, et al. 2013. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**: 498–503. doi:10.1038/nature12111
- Howe K, Schiffer PH, Zielinski J, Wiehe T, Laird GK, Marioni JC, Soylemez O, Kondrashov F, Leptin M. 2016. Structure and evolutionary history of a large family of NLR proteins in the zebrafish. *Open Biol* **6**: 160009. doi:10.1098/rsob.160009
- Huang W, Li L, Myers JR, Marth GT. 2012. ART: a next-generation sequencing read simulator. *Bioinformatics* **28**: 593–594. doi:10.1093/bioinformatics/btr708
- Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, Arcila D, Betancur-R, Li C, Becker L, et al. 2018. Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. *Proc Natl Acad Sci* **115**: 6249–6254. doi:10.1073/pnas.1719358115
- Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L. 2006. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res* **16**: 669–677. doi:10.1101/gr.4842106
- Imbeault M, Hellebood P-Y, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**: 550–554. doi:10.1038/nature21683
- Jacobs FMJ, Greenberg D, Nguyen N, Haussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D. 2014. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516**: 242–245. doi:10.1038/nature13760

- Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV. 2004. Duplication, coclustering, and selection of human Alu retrotransposons. *Proc Natl Acad Sci* **101**: 1268–1272. doi:10.1073/pnas.0308084100
- Kaplan T, Friedman N, Margalit H. 2005. Ab initio prediction of transcription factor targets using structural knowledge. *PLoS Comput Biol* **1**: e1. doi:10.1371/journal.pcbi.0010001
- Kasinathan B, Colmenares SU, McConnell H, Young JM, Karpen GH, Malik HS. 2020. Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in *Drosophila*. *eLife* **9**: e63368. doi:10.7554/eLife.63368
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780. doi:10.1093/molbev/mst010
- Keller O, Kollmar M, Stanke N, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* **27**: 757–763. doi:10.1093/bioinformatics/btr010
- Kent TV, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philos Trans R Soc Lond B Biol Sci* **372**: 20160458. doi:10.1098/rstb.2016.0458
- Klug A. 2010. The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation. *Q Rev Biophys* **43**: 1–21. doi:10.1017/S0033583510000089
- Kofler R, Betancourt AJ, Schlötterer C. 2012. Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in *Drosophila melanogaster*. *PLoS Genet* **8**: e1002487. doi:10.1371/journal.pgen.1002487
- Kosiol C, Holmes I, Goldman N. 2007. An empirical codon model for protein sequence evolution. *Mol Biol Evol* **24**: 1464–1479. doi:10.1093/molbev/msm064
- Kumar S, Suleski M, Craig JM, Kasprowitz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: an expanded resource for species divergence times. *Mol Biol Evol* **39**: msac174. doi:10.1093/molbev/msac174
- Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330. doi:10.1038/nature14248
- Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. 1988. On the role of unequal exchange in the containment of transposable element copy number. *Genet Res* **52**: 223–235. doi:10.1017/S0016672300027695
- Laue K, Rajshekar S, Courtney AJ, Lewis ZA, Goll MG. 2019. The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo. *Nat Commun* **10**: 1551. doi:10.1038/s41467-019-09582-3
- Li HD, Lin CX, Zheng J. 2022. GTFtools: a software package for analyzing various features of gene models. *Bioinformatics* **38**: 4806–4808. doi:10.1093/bioinformatics/btac561
- Liu H, Chang L-H, Sun Y, Lu X, Stubbs L. 2014. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. *Genome Biol Evol* **6**: 510–525. doi:10.1093/gbe/evu030
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. *Methods Mol Biol* **1079**: 155–170. doi:10.1007/978-1-62703-646-7_10
- Margalit L, Strauss C, Tal A, Schlesinger S. 2020. Trim24 and Trim33 play a role in epigenetic silencing of retroviruses in embryonic stem cells. *Viruses* **12**: 1015. doi:10.3390/v12091015
- Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y. 2010. Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**: 927–931. doi:10.1038/nature08858
- McLaughlin RN, Malik HS. 2017. Genetic conflicts: the usual suspects and beyond. *J Exp Biol* **220**: 6–17. doi:10.1242/jeb.148148
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* **37**: 1530–1534. doi:10.1093/molbev/msaa015
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res* **49**: D412–D419. doi:10.1093/nar/gkaa913
- Molparia B, Goyal K, Sarkar A, Kumar S, Sundar D. 2010. ZiF-predict: a web tool for predicting DNA-binding specificity in C2H2 zinc finger proteins. *Genomics Proteomics Bioinformatics* **8**: 122–126. doi:10.1016/S1672-0229(10)60013-7
- Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM, et al. 2015. C2h2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat Biotechnol* **33**: 555–562. doi:10.1038/nbt.3128
- Najafabadi HS, Garton M, Weirauch MT, Mnaimneh S, Yang A, Kim PM, Hughes TR. 2017. Non-base-contacting residues enable kaleidoscopic evolution of metazoan C2H2 zinc finger DNA binding. *Genome Biol* **18**: 167. doi:10.1186/s13059-017-1287-y
- Nowick K, Fields C, Gernat T, Caetano-Anolles D, Kholina N, Stubbs L. 2011. Gain, loss and divergence in primate zinc-finger genes: a rich resource for evolution of gene regulatory differences between species. *PLoS One* **6**: e21553. doi:10.1371/journal.pone.0021553
- Panfili KA, Vargas Jentsch IM, Benoit JB, Erezylmaz D, Suzuki Y, Colella S, Robertson HM, Poelchau MF, Waterhouse RM, Ioannidis P, et al. 2019. Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *Genome Biol* **20**: 64. doi:10.1186/s13059-019-1660-0
- Panyushev N, Okorokova L, Danilov L, Adonin L. 2021. Pattern of repetitive element transcription segregate cell lineages during the embryogenesis of sea urchin *Strongylocentrotus purpuratus*. *Biomedicines* **9**: 1736. doi:10.3390/biomedicines9111736
- Parisot N, Vargas-Chávez C, Goubert C, Baa-Puyoulet P, Balmand S, Beranger L, Blanc C, Bonnamour A, Boulesteix M, Burlet N, et al. 2021. The transposable element-rich genome of the cereal pest *Sitophilus oryzae*. *BMC Biol* **19**: 241. doi:10.1186/s12915-021-01158-2
- Persikov AV, Singh M. 2014. De novo prediction of DNA-binding specificities for Cys₂His₂ zinc finger proteins. *Nucleic Acids Res* **42**: 97–108. doi:10.1093/nar/gkt890
- Petrov DA. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol* **20**: 880–892. doi:10.1093/molbev/msg102
- Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, Theunissen TW, Jaenisch R, Trono D. 2019. Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* **24**: 724–735.e5. doi:10.1016/j.stem.2019.03.012
- Pontis J, Pulver C, Playfoot CJ, Planet E, Grun D, Offner S, Duc J, Manfrin A, Lutolf MP, Trono D. 2022. Primate-specific transposable elements shape transcriptional networks during human development. *Nat Commun* **13**: 7178. doi:10.1038/s41467-022-34800-w
- Potireddy S, Vassena R, Patel BG, Latham KE. 2006. Analysis of polysomal mRNA populations of mouse oocytes and zygotes: dynamic changes in maternal mRNA utilization and function. *Dev Biol* **298**: 155–166. doi:10.1016/j.ydbio.2006.06.024
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet* **16**: 276–277. doi:10.1016/S0168-9525(00)02024-2
- Rodriguez-Terrones D, Torres-Padilla M-E. 2018. Nimble and ready to mingle: transposon outbursts of early development. *Trends Genet* **34**: 806–820. doi:10.1016/j.tig.2018.06.006
- Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, et al. 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**: 237–240. doi:10.1038/nature08674
- Sasaki M, Lange J, Keeney S. 2010. Genome destabilization by homologous recombination in the germ line. *Nat Rev Mol Cell Biol* **11**: 182–195. doi:10.1038/nrm2849
- Seroussi U, Lugowski A, Wadi L, Lao RX, Willis AR, Zhao W, Sundby AE, Charlesworth AG, Reinke AW, Claycomb JM. 2023. A comprehensive survey of *C. elegans* argonaute proteins reveals organism-wide gene regulatory networks and functions. *eLife* **12**: e83853. doi:10.7554/eLife.83853
- Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **151**: 964–980. doi:10.1016/j.cell.2012.10.040
- Sniegowski PD, Murphy HA. 2006. Evolvability. *Curr Biol* **16**: R831–R834. doi:10.1016/j.cub.2006.08.080
- Song M, Boissinot S. 2007. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. *Gene* **390**: 206–213. doi:10.1016/j.gene.2006.09.033
- Tadepally HD, Burger G, Aubry M. 2008. Evolution of C2H2-zinc finger genes and subfamilies in mammals: species-specific duplication and loss of clusters, genes and effector domains. *BMC Evol Biol* **8**: 176. doi:10.1186/1471-2148-8-176
- Tarlington RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. *Nature* **442**: 79–81. doi:10.1038/nature04841
- Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Res* **21**: 1800–1812. doi:10.1101/gr.121749.111
- To TH, Jung M, Lycett S, Gascuel O. 2016. Fast dating using least-squares criteria and algorithms. *Syst Biol* **65**: 82–97. doi:10.1093/sysbio/syv068

- Tollis M, Boissinot S. 2012. The evolutionary dynamics of transposable elements in eukaryote genomes. *Genome Dyn* **7**: 68–91. doi:10.1159/000337126
- Vacic V, Uversky VN, Dunker AK, Lonardi S. 2007. Composition profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics* **8**: 211. doi:10.1186/1471-2105-8-211
- Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**: 1833–1837. doi:10.1126/science.1074973
- Wei KH-C, Chan C, Bachtrog D. 2021. Establishment of H3K9me3-dependent heterochromatin during embryogenesis in *Drosophila miranda*. *eLife* **10**: e55612. doi:10.7554/eLife.55612
- White RJ, Collins JE, Sealy IM, Wali N, Dooley CM, Digby Z, Stemple DL, Murphy DN, Billis K, Hourlier T, et al. 2017. A high-resolution mRNA expression time course of embryonic development in zebrafish. *eLife* **6**: e30860. doi:10.7554/eLife.30860
- Winata CL, Łapiński M, Prysycz L, Vaz C, bin Ismail MH, Nama S, Hajan HS, Lee SG, Korzh V, Sampath P, et al. 2018. Cytoplasmic polyadenylation-mediated translational control of maternal mRNAs directs maternal-to-zygotic transition. *Development* **145**: dev159566. doi:10.1242/dev.159566
- Wolf D, Goff SP. 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**: 1201–1204. doi:10.1038/nature07844
- Wolf G, Yang P, Füchtbauer AC, Füchtbauer E-M, Silva AM, Park C, Wu W, Nielsen AL, Pedersen FS, Macfarlan TS. 2015. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev* **29**: 538–554. doi:10.1101/gad.252767.114
- Wolf G, de Iaco A, Sun MA, Bruno M, Tinkham M, Hoang D, Mitra A, Ralls S, Trono D, Macfarlan TS. 2020. KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *eLife* **9**: e56337. doi:10.7554/eLife.56337
- Wolfe SA, Nekudova L, Pabo CO. 2000. DNA recognition by Cys₂His₂ zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**: 183–212. doi:10.1146/annurev.biophys.29.1.183
- Wright SI, Agrawal N, Bureau TE. 2003. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Res* **13**: 1897–1903. doi:10.1101/gr.1281503
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591. doi:10.1093/molbev/msm088
- Yu T, Koppetsch BS, Pagliarani S, Johnston S, Silverstein NJ, Luban J, Chappell K, Weng Z, Theurkauf WE. 2019. The piRNA response to retroviral invasion of the koala genome. *Cell* **179**: 632–643.e12. doi:10.1016/j.cell.2019.09.002

Received April 7, 2023; accepted in revised form June 15, 2023.