

Research Article

Genetic Innovation in Vertebrates: Gypsy Integrase Genes and Other Genes Derived from Transposable Elements

Domitille Chalopin, Delphine Galiana, and Jean-Nicolas Volff

Institut de Génomique Fonctionnelle de Lyon, Université de Lyon, Ecole Normale Supérieure de Lyon, CNRS, Université Lyon 1, 69364 Lyon Cedex 07, France

Correspondence should be addressed to Domitille Chalopin, domitille.chalopin@ens-lyon.fr

Received 25 May 2012; Accepted 15 July 2012

Academic Editor: Wen Wang

Copyright © 2012 Domitille Chalopin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to their ability to drive DNA rearrangements and to serve as a source of new coding and regulatory sequences, transposable elements (TEs) are considered as powerful evolutionary agents within genomes. In this paper, we review the mechanism of molecular domestication, which corresponds to the formation of new genes derived from TE sequences. Many genes derived from retroelements and DNA transposons have been identified in mammals and other vertebrates, some of them fulfilling essential functions for the development and survival of their host organisms. We will particularly focus on the evolution and expression of Gypsy integrase (*GIN*) genes, which have been formed from ancient event(s) of molecular domestication and have evolved differentially in some vertebrate sublineages. What we describe here is probably only the tip of the evolutionary iceberg, and future genome analyses will certainly uncover new TE-derived genes and biological functions driving genetic innovation in vertebrates and other organisms.

1. Introduction

For a long time, transposable elements (TEs) have been considered as pure selfish and junk elements parasiting the genome of living organism [1, 2]. These sequences are able to “move”, that is, to insert into new locations within genomes. This phenomenon is called transposition. Retroelements use retrotransposition, that is, the reverse transcription of an RNA intermediate and integration of the cDNA molecule produced, to generate new copies of themselves within genomes (copy-and-paste mechanism). This mechanism directly increases the copy number of the element. Among protein-coding autonomous retroelements, distinction is generally made between elements with long terminal repeats (LTRs: LTR retrotransposons and retroviruses) and retroelements without LTRs (non-LTR retrotransposons or LINE elements). Retroviruses and LTR retrotransposons are mainly distinguished by the presence versus absence of an envelope gene, which encodes a protein necessary for virus entry into the target cell. After germ line infection, reverse-transcribed retrovirus genomes can be

integrated into the host genome and transmitted through vertical inheritance to the host progeny [3]. Such sequences, called endogenous retroviruses, are generally inactivated by mutations. Gain or loss of the envelope gene can transform a retrotransposon into a retrovirus, and *vice versa* [4, 5]. The second large category of TEs, DNA transposons, generally excises from their original insertion site and reintegrate into a new location (cut-and-paste mechanism). For most DNA transposons, transposition is catalyzed by an enzyme called transposase [6]. Finally, noncoding nonautonomous elements using for their transposition proteins encoded by autonomous sequences exist for both retroelements and DNA transposons.

Despite the deep-rooted vision of junk DNA, there is growing evidence that TEs are more than simple genome parasites. Particularly, they have been shown to serve as a genomic reservoir for new regulatory and coding sequences allowing genetic innovation and organismal evolution. A fascinating facet of the roles of TEs in evolution is their ability to be “molecularly domesticated” to form new cellular protein-coding genes [7, 8]. TE-encoded proteins have properties

that can be of interest for host cellular pathways. They can bind, copy, cut, process, and recombine nucleic acids, as well as modify and interact with host proteins. There are many cases of TE-derived genes fulfilling important functions in plants, fungi, and animals, including vertebrates (for review, [8, 9]). We will present here several prominent examples of vertebrate genes formed from TE-coding sequences during evolution, with more emphasis on Gypsy integrase (*GIN*) genes that we have analyzed in different fish species.

2. Genes Derived from Retroelements

2.1. Gag-Derived Genes. Several multigenic families have been formed from different events of molecular domestication of the *gag* gene of Ty3/Gypsy elements, a super family of LTR retrotransposons active in fish and amphibians but extinct in mammals [9, 10]. The *gag* gene encodes a structural protein with three functional regions: the matrix (MA) domain playing a role in targeting cellular membranes, the capsid (CA) domain involved in interactions with other proteins during particle assembly, and the nucleocapsid (NC), which binds to viral RNA genomes through zinc fingers.

One *gag*-related gene family is called *Mart*. This gene family is mammal specific and constituted by 12 genes in human [11]. Most *Mart* genes are found on mammalian X chromosome, suggesting an initial event of molecular domestication on the X, followed by serial local duplication events that subsequently extended this gene family. All *Mart* genes have retained from the original *gag* sequence an intronless open reading frame. Some of them still encode the ancestral Gag zinc finger, suggesting nucleic acid binding properties for the protein. Two autosomal *Mart* genes, *PEG10 (Mart2)* and *PEG11/Rtl1 (Mart1)*, are subject to genomic imprinting and are expressed from the paternal allele [12, 13]. This epigenetic regulation has been proposed to be derived from a defence mechanism repressing the activity of the ancestral retrotransposon before domestication [14]. At least two *Mart* genes, *PEG11/Rtl1 (Mart1)* and *PEG10 (Mart2)*, have essential but nonredundant functions in placenta development in the mouse [15, 16]. *PEG10* and other *Mart* genes might also control cell proliferation and apoptosis, with possible involvement in cancer ([8] and references therein).

Another mammalian gene family derived from a LTR retrotransposon *gag* gene is called *Ma* or *Pnma* (paraneoplastic Ma antigens) [17]. Fifteen *Ma/Pnma* genes are present in the human genome, most of them being located on the X chromosome as observed for *Mart* genes. Some Ma proteins are expressed by patients with paraneoplastic neurological disorders and might be targeted by autoimmune response leading to progressive neurological damage [18]. Several Ma proteins are also involved in apoptosis, including Ma4 (*Pnma4/Map1/Maop1*) and Ma1/*Pnma1* [19, 20].

A third family is the SCAN domain family. This family is constituted of DNA binding proteins with an N-terminus region called the SCAN domain, which is derived from the Gag protein of a Gmr1-like Gypsy/Ty3 retrotransposon

[21–24]. The SCAN family is vertebrate specific, with approximately 70 and 40 members in human and mouse, respectively. Several SCAN proteins have been shown to be transcription factors regulating diverse biological processes such as hematopoiesis, stem cell properties, or cell proliferation and apoptosis (for review [21]).

Finally, other *gag*-related genes are present in mammalian genomes [10]. One of them, *Fv1*, is of retroviral origin and controls replication of the murine leukaemia virus in the mouse [25].

2.2. Envelope-Derived Genes. During mammalian evolution, retroviral envelope genes have been domesticated several times independently to generate genes involved in placenta development [26]. These genes, derived from endogenous retroviruses, encode proteins called syncytins. Syncytins mediate the fusion of trophoblast cells to form the syncytiotrophoblast layer, a continuous structure with microvillar surfaces forming the outermost foetal component of the placenta [27]. Two syncytin genes of independent origins encoding placenta-specific fusogenic proteins are present in human and other simians (*Syncytin-1* and *-2*, [28]) as well as in rodents (*Syncytin-A* and *Syncytin-B*, [29]). Independent *Syncytin* genes are also found in rabbit [30], guinea pig [31], and Carnivora [32], indicating multiple convergent domestication of *env*-derived *Syncytin* genes in different mammalian sublineages. Some *Syncytins* might be involved in other biological processes. For example, human *Syncytin-1* plays a role in osteoclast fusion, neuroinflammation, and possibly multiple sclerosis [33, 34].

Other retroviral *env*-derived open reading frames are present in vertebrate genomes; but intensive work is required to determine their functions. Some of them might confer resistance to viral infection, as shown for the *Fv-4* locus. This locus, containing an entire ecotropic murine leukemia virus (MuLV) *env* gene, controls susceptibility to infection by MuLV [35].

2.3. Other Retroelement-Derived Genes. In mammals, a gene called *CGIN1* is partially derived from the integrase gene of an endogenous retrovirus. The integrase gene has been fused 125–180 million years ago to a duplicate of the cellular gene *KIAA0323*. A role of *CGIN1* in resistance against retroviruses has been proposed [36].

Several genes with homology to retroelement aspartyl protease genes are present in vertebrate genomes. One of them, a gene encoding a protein called SASPase, is necessary for the texture and hydration of the stratum corneum, the outermost layer of the epidermis [37].

Finally, the telomerase, the reverse transcriptase extending the ends of linear chromosomes in vertebrates and other eukaryotes, might be derived from a retroelement [38].

3. Genes Derived from DNA Transposons

Many examples of genes derived from transposase genes from diverse subfamilies of DNA transposons have been described in vertebrates and other organisms [8, 39, 40].

One well-studied example is the recombination-activating protein Rag1, which together with Rag2 catalyzes the V(D)J somatic site-specific recombination responsible for the formation and diversity of genes encoding immunoglobulins and T-cell receptors in jawed vertebrates. *Rag1* has been formed from the transposase of a Transib DNA transposon, and the V(D)J recombination signal sequences recognized by Rag1 might be derived from the transposon ends bound by the ancestral transposase [41].

The mammal-specific gene *CENP-B* encodes a Pogo transposase-derived protein that controls centromere formation depending on the chromatin context [42]. Interestingly, an independent event of molecular domestication of Pogo transposase also led to the formation of centromeric proteins in fission yeast [43]. In yeast, CENP-B-like proteins restrict the activity of retrotransposons and promote replication progression at forks paused by retrotransposon LTRs [44, 45]. Other genes are derived from Pogo-like transposons in mammals [46]. One example is the *Jerky* gene, which encodes a brain-specific mRNA-binding protein that may regulate mRNA use in neurons [47].

Similarly, several examples of genes derived from hAT transposases have been found in mammals, some of them having been fused to zinc finger domains [46]. Some hAT transposase-related proteins work as transcription factors. One of them, ZEBD6/MGR, negatively regulates *IGF2* expression and muscle growth. Indeed, it has been shown that mutation in a regulatory sequence prohibiting ZEBD6/MGR binding leads to *IGF2* upregulation and enhanced muscle growth in commercially bred pigs [48, 49].

In primates, the gene encoding the Metnase/SETMAR protein has been formed through fusion of the transposase gene of a Mariner transposon with a SET histone methyltransferase gene. Metnase/SETMAR is a DNA binding protein with endonuclease activity that promotes DNA double-strand break repair through nonhomologous end joining (NHEJ) [50, 51].

Several genes derived from PiggyBac-like transposons have been detected in human and other vertebrates [52]. One of them, *PGBD3*, serves as an alternative 3' terminal exon for the Cockayne Syndrome B (CSB) gene, leading to the expression of a CSB-transposase fusion protein [53]. At least one Harbinger transposon-derived gene, *HARB1*, encoding a predicted nuclease, is present in mammals, birds, amphibians, and fish [54]. Likewise, genes derived from a new type of DNA transposon called Zisupton have been identified in fish and other vertebrates [55]. Finally, mammalian and bird genomes possess at least one gene clearly derived from a P transposon; additional vertebrate genes like *THAP9* encoding proteins with a THAP domain might be also related to P-like transposases [8, 40, 56–61].

4. Gypsy Integrase Genes: Data from Fish

Two vertebrate genes with unknown functions, *GIN1* and *GIN2* (*Gypsy Integrase 1 and 2*), encode proteins showing significant homologies to integrases encoded by LTR retrotransposons [62, 63]. Further analyses showed that both genes have been formed from GIN transposons, a new family

of metazoan DNA transposons with a transposase that shows strong similarities with LTR retrotransposon integrases [64]. *GIN1*, which shows similarities with GINO transposons from *Hydra magnipapillata*, is present in mammals, birds, and reptiles, suggesting a molecular domestication event at the base of the Amniota ca. 300 million years ago. Mammalian *GIN1* proteins have conserved amino-acid residues necessary for integrase activity. Using our own analyses, we will now particularly focus on the *GIN2* gene. We provide here updated *GIN2* structural and phylogenetic analyses using new vertebrate sequences and present first expression data for this gene in fish.

GIN2 is present in several fish species, as well as in cartilaginous fish (elephant shark), coelacanth, amphibians, birds, reptiles, and marsupials, but neither in monotremes nor in placental mammals [63] (Figures 1, 2, and 3). Furthermore, *GIN2* was not detected in lamprey. Hence, the molecular domestication event having led to the formation of *GIN2* might have taken place before the divergence between tetrapods/bony fish and cartilaginous fish around 500 million years ago, with subsequent loss in monotremes and placental mammals. The formation of *GIN2* might even be older, since potentially domesticated GIN-like sequences related to *GIN2* have been detected in the urochordates *Ciona savignyi* and *C. intestinalis* [63]. Phylogenetic analysis suggests that *GIN2* is derived from GINA transposons, which are *bona fide* transposable elements in *Hydra magnipapillata* (Figure 1). This suggests that *GIN1* and *GIN2* have been formed through two independent molecular domestication events, one at the base of Amniota and the other in a more ancient vertebrate ancestor (Figure 3).

After domestication, the HHCC zinc finger present in the ancestral integrase has been maintained, suggesting ability to bind to DNA or RNA (Figure 2). Conservation of the important catalytic triad (DDE, aspartic acid/aspartic acid/glutamic acid) of the integrase is less obvious. While this motif has been proposed to be conserved in *GIN1*, this is not the case for *GIN2* based on a published alignment with sequences from GIN-related transposases [63] (Figure 2). As shown in Figure 2, the first aspartic acid residue is present in most species but absent from amphibians and birds. However, multiple sequence alignment revealed an aspartate conserved in all *GIN2* and *GIN1* sequences ca. 20 amino-acids downstream. The second aspartic acid residue is not found in *GIN2* but an aspartate is conserved four amino acids away in all *GIN2* sequences except for opossum. Finally, the glutamic acid residue is found only in several species and substituted by an aspartate in fish; but a conserved glutamate is detected 16 amino acids away. Hence, the question of the functionality of *GIN2* as an integrase remains open and should be definitely answered through functional analyses. A third domain with unknown function called GPY/F [64, 67] is also detected in GIN proteins, but in some cases the phenylalanine residue is replaced by a leucine. *GIN2* contains eight protein-coding exons, with an exon-intron structure well conserved in fish and other vertebrates (Figure 4). Some introns might be derived from the ancestral transposon; others might be the result of events of intronization after molecular domestication. *GIN2* is located in the same

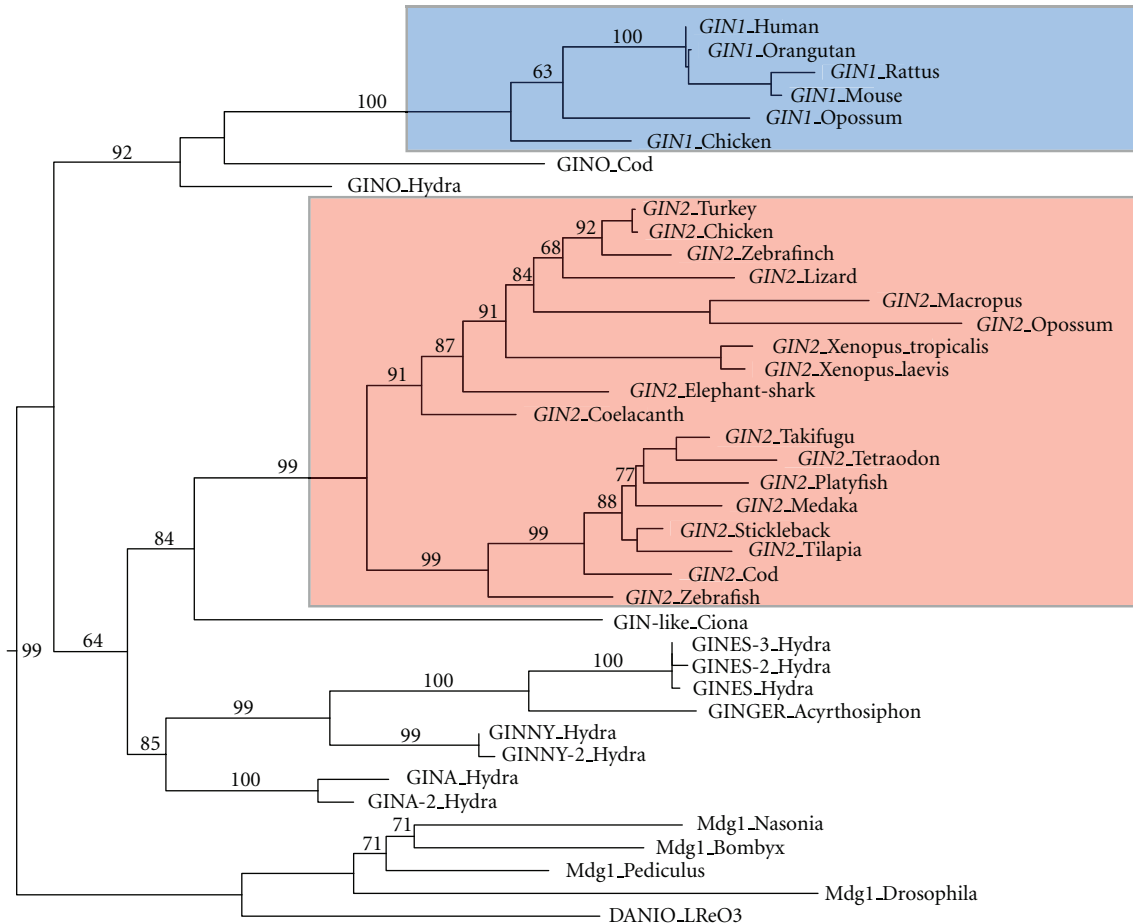


FIGURE 1: Molecular phylogeny of GIN proteins. Phylogenetic tree based on a 352 amino-acid integrase alignment. Protein sequences were aligned with clustalW and phylogenetic tree was constructed using maximum likelihood from PhyML package (optimized default bootstrap) [65]. Sequences were recovered from NCBI and Ensembl or predicted from genome sequences. Accession numbers and sequence alignments are available upon request.

orthologous genomic region between *OGFOD2* and *ABCB9* in marsupials, birds, reptiles, and fish, confirming that this gene does not correspond to a mobile sequence (Figure 5).

Expressed sequence tag (EST) analysis indicated that *GIN2* is expressed in different adult tissues and developmental stages in chicken: brain (accession number: CN219658), liver (BG713188), head (BU225420), embryonic tissue (BU210425), limb (BU256599), small intestine (BU297502), muscle (BU437928), and ovary (BU447634). Only ESTs from the whole body are available for *Xenopus*. Few ESTs are also found in zebrafish: muscle (CT684014), gills (EB908574), reproductive system (BI867074), and eye (BI879358).

To determine more precisely *GIN2* expression pattern in fish, quantitative real-time PCR was performed on different embryonic developmental stages in zebrafish (*Danio rerio*), as well as on adult tissues from zebrafish and platyfish (*Xiphophorus maculatus*) (Figure 6). During zebrafish embryogenesis, *GIN2* expression level strongly increases from the dome stage and progressively decreases until the end of somite stages. This result suggests that *GIN2* possibly plays a role during gastrulation. Gastrulation, which is

characterized by morphologic movements of involution and extension, starts at the beginning of the epiboly to finish at bud stage [68]. In adult zebrafish, the higher level of expression for *GIN2* was observed in brain, followed by gonads and eyes. In contrast, *GIN2* expression was maximal in gonads in the platyfish (Figure 6).

To conclude, our analysis integrates data from several newly sequenced vertebrate genomes, particularly teleostean and cartilaginous fishes as well as coelacanth, in order to better understand the distribution and evolutionary history of *GIN* genes. Since *GIN2* is apparently not present in lamprey, we propose that *GIN2* was formed before the divergence between cartilaginous and ray-finned fish about 500 million years ago (Figure 3). We also provide the first expression data for *GIN2* in fish particularly supporting a function in gastrulation during zebrafish embryogenesis.

5. Conclusion

At first glance, transposable elements were considered as “junk” DNA with no important functions for genomes and

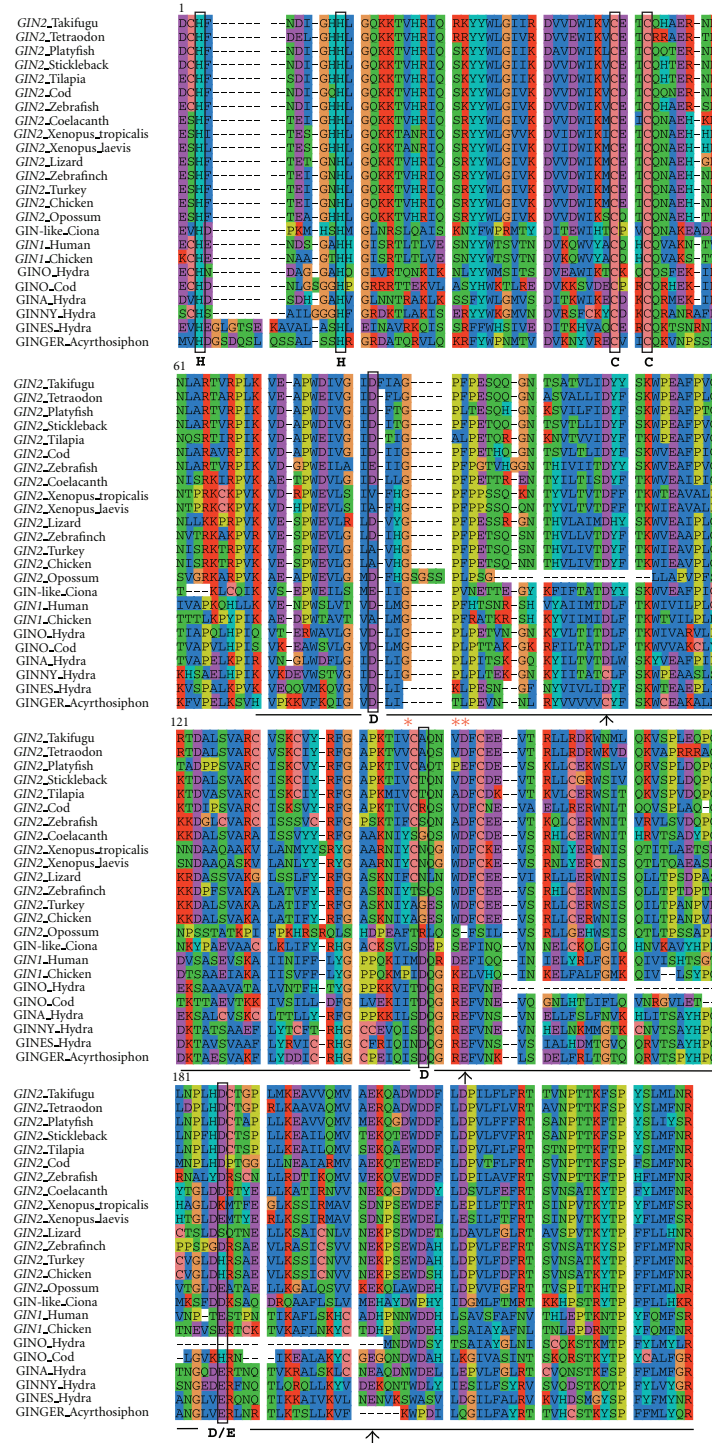


FIGURE 2: Sequence alignment of predicted GIN2-related proteins. HHCC zinc finger and integrase-like domain of GIN2 were aligned using clustalW [66]. The black line indicates the position of the integrase-like domain. HHCC and DDE motifs are shown in black boxes and GFP motif is highlighted by red asterisks. Arrows indicate alternative conserved D/E residues in GIN2 sequences.

organisms. Today, nobody can deny the importance of transposable elements during evolution in terms of innovation power, particularly through molecular domestication events. Domesticated elements are *bona fide* cellular genes derived from transposable element sequences encoding for example

integrase, transposases, Gag proteins, or envelopes. After domestication, TE-derived genes have lost their ability to transpose through the elimination of sequences such as long terminal repeats, terminal-inverted repeats, or other open reading frames and protein domains essential for

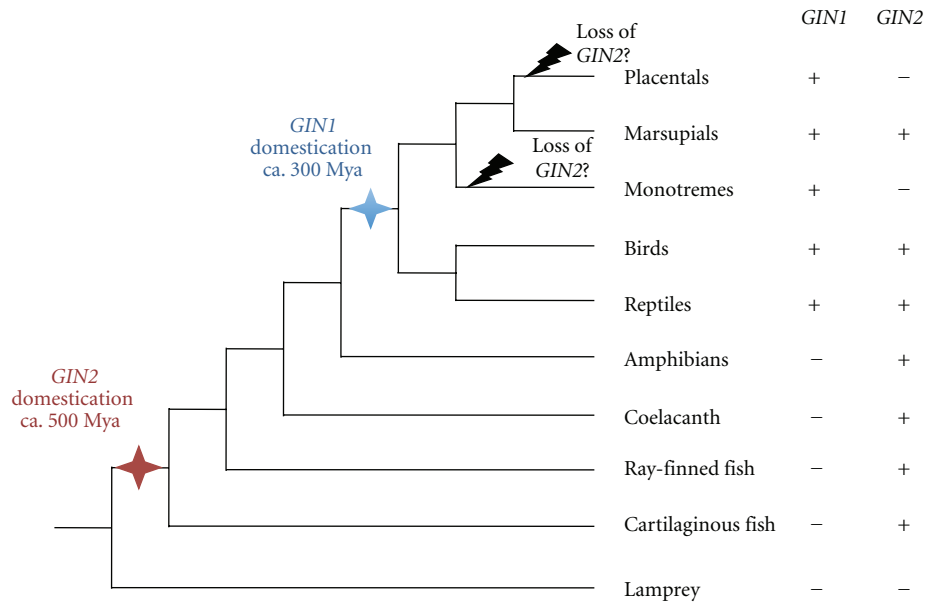


FIGURE 3: One possible scenario for the evolution of *GIN* genes in vertebrates. The two molecular domestication events are highlighted by blue and red stars, having led to the formation of *GIN1* and *GIN2*, respectively. Presence (+) or absence (-) of *GIN1* and *GIN2* in the different lineages is indicated.

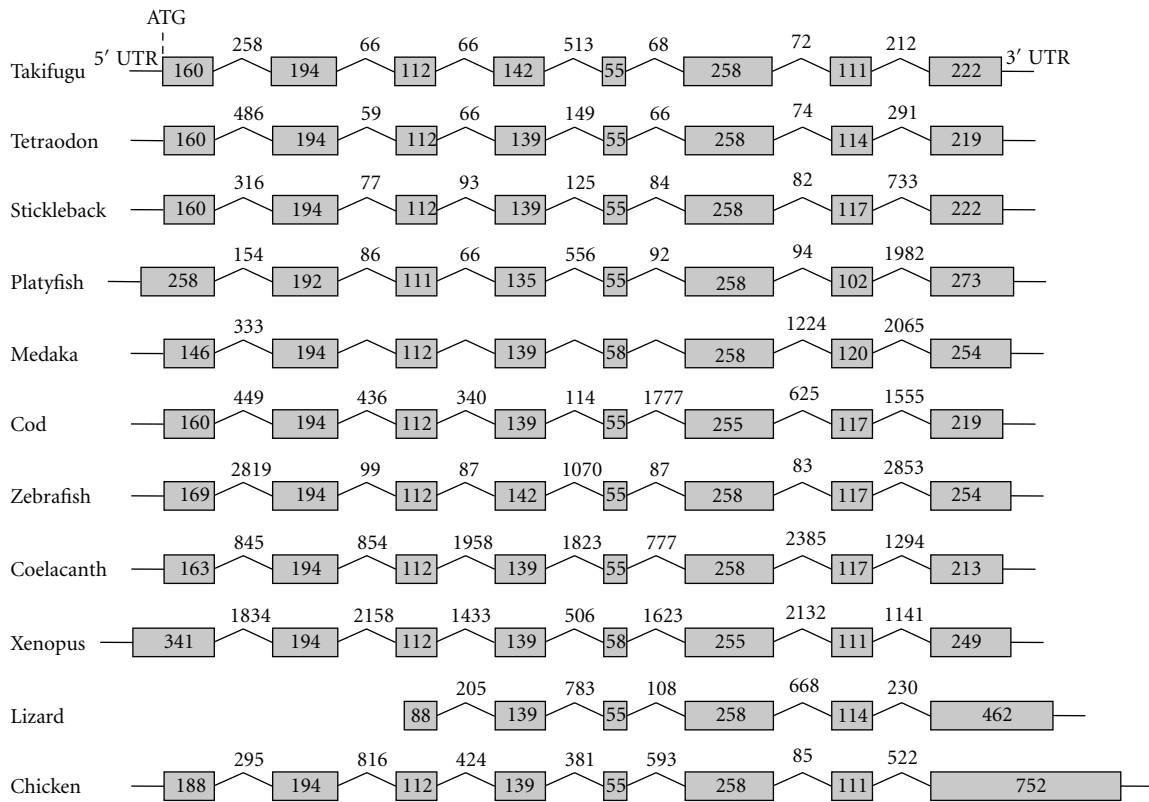


FIGURE 4: Exon-intron structure of *GIN2* genes in fish and other vertebrates. Exons are represented by grey boxes, introns by broken lines. Exon/intron sizes are given as base pairs.

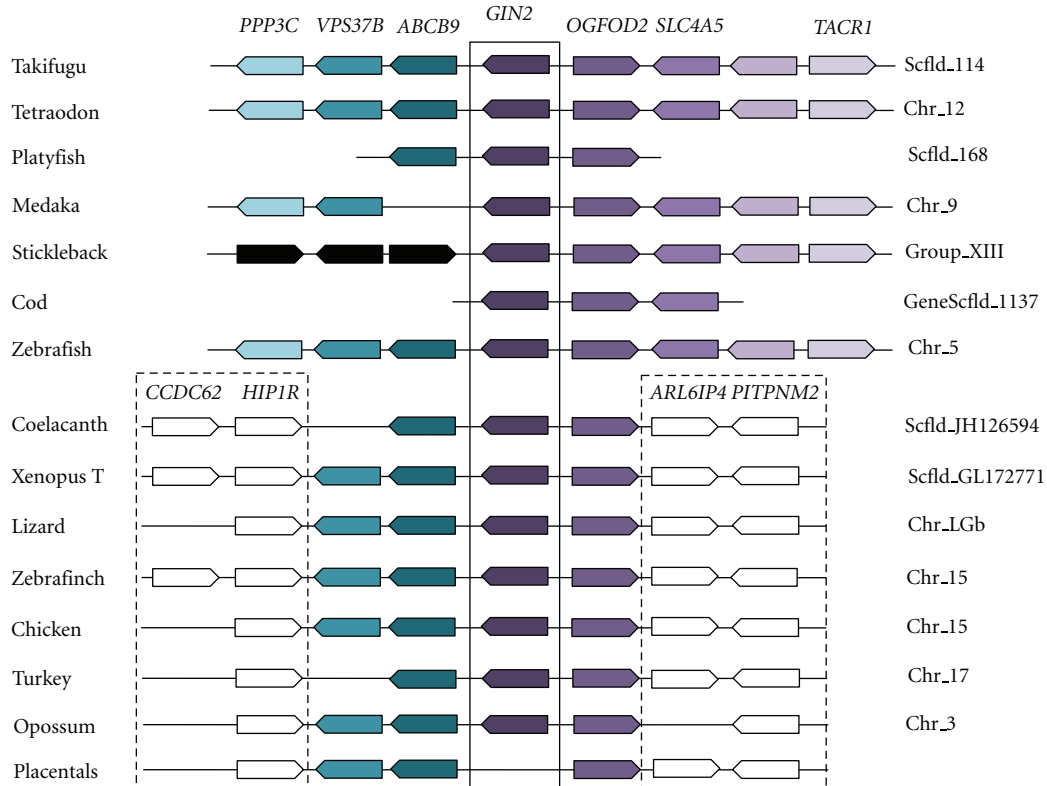


FIGURE 5: Comparison of *GIN2*-containing genomic regions in vertebrates. Synteny analysis was performed using Ensembl (<http://www.ensembl.org/index.html>), Genomicus (<http://www.dyogen.ens.fr/genomicus-66.01/cgi-bin/search.pl>), and BLAST analysis (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Genes represented by white boxes are not found in this region in fish. Black boxes from stickleback represent different genes from *ABCB9*, *VPS37B*, and *PPP3C* in other fish. Their Ensembl accession numbers are from right to left: ENSGACG00000013652, ENSGACG00000013659, and ENSGACG13660.

transposition. Elimination of such sequences might occur by genetic drift or might even be selected for transposition or retrotransposition of a domesticated sequence might change its copy number and pattern of expression. Many domesticated sequences have important functions, for example in cell proliferation. Transposition of such a gene might have strongly deleterious consequences for the host, for instance cancer. It might, therefore, be important to immobilize TE-derived genes at fixed position within a genome to control their expression.

In vertebrates, many TE-derived genes are mammal specific, suggesting that molecular domestication probably played an important role in the evolution of this specific sublineage. Accordingly, many domesticated sequences are involved in placenta formation. Other TE-derived genes like *GIN2* are present in some vertebrate sublineages but absent from mammals. In birds, reptiles, amphibians, and fish, domesticated sequences might be more difficult to identify due to the concomitant presence of active TEs within genomes. Availability of additional genome sequences will probably allow the identification of many TE-derived genes specific of these sublineages that contribute to diversification within vertebrates.

We focused on *GIN* genes, a pair of ancient vertebrate domesticated genes for which no function has been identified so far. Both *GIN1* and *GIN2* are derived from *GIN* transposons that themselves gained their transposase from the integrase of LTR retrotransposons.

GIN1 was detected in mammals, birds, and reptiles, indicating that it was formed in a common ancestor of Amniota ca. 300 million years ago [63]. *GIN2* might be even older, since it was detected in tetrapods, bony fish, and sharks, and possibly in urochordates. The presence of both genes over such long periods of evolution is suggestive of important, so far unknown conserved functions in vertebrates. *GIN2* was lost in a common ancestor of monotremes and placental mammals, suggesting that either *GIN2* function was not essential anymore, or that this function is fulfilled now by *GIN1* in these sublineages.

The evolutionary scenario having led to the formation of *GIN1* and *GIN2* remains unclear. Presence of conserved intron positions [63] suggests a unique origin followed by duplication and intron gain in a common ancestor of *GIN1* and *GIN2* (paralogy). In this case, *GIN1* would have been lost among others in fish. Alternatively, *GIN1* and *GIN2* might have been generated from two independent

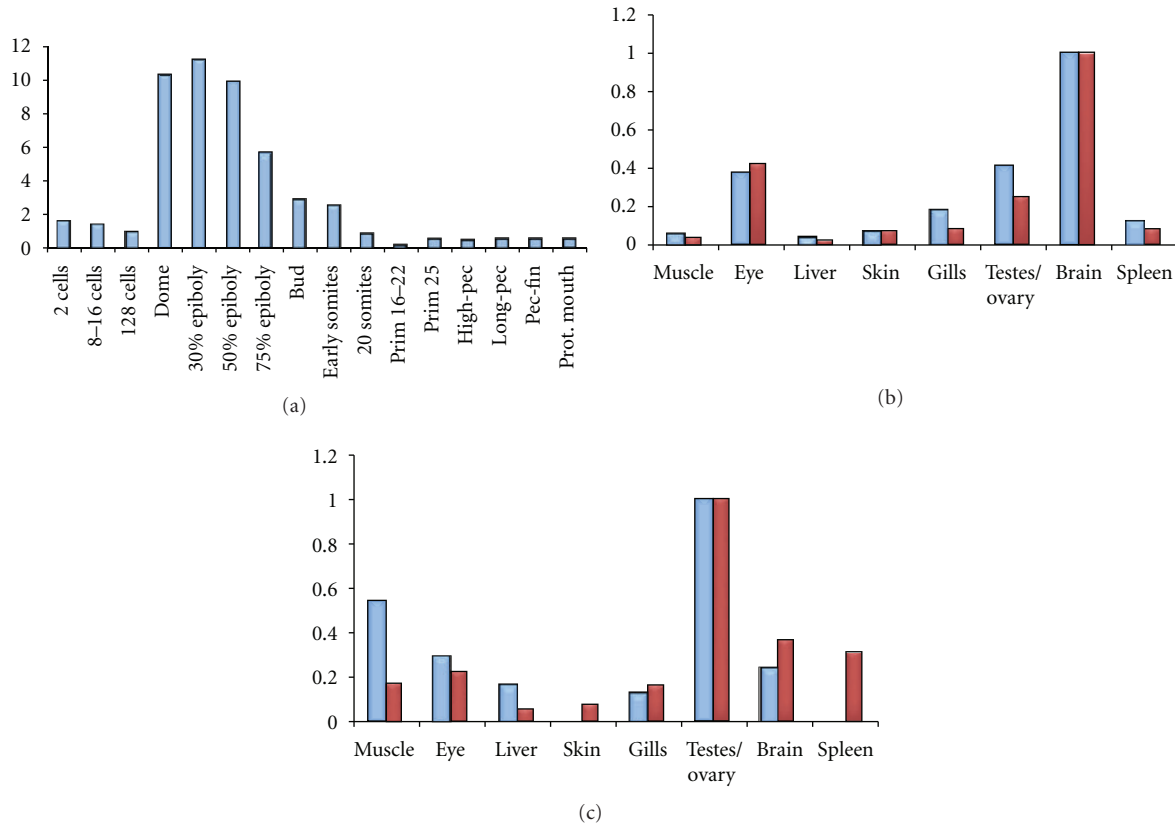


FIGURE 6: qPCR expression analysis of *GIN2* in zebrafish and platyfish. (a) Expression pattern of *GIN2* during embryonic development in zebrafish. (b) Expression pattern of *GIN2* in adult organs of zebrafish. (c) Expression pattern of *GIN2* in adult organs of platyfish. Multiple RNA extractions using different individuals were performed leading to independent sets of cDNA. Two independent sets and three independent sets of cDNA were tested for embryonic stages and adult organs, respectively. For all sets and for each sample of cDNA, qPCR reaction was done three times (triplicate). One representative experiment is shown with blue bars for male samples and red bars for female samples. *GIN2* expression was normalized using three housekeeping genes: *RPL7*, *beta-actin* and *EF1-alpha*. Analyses were done using the $\Delta\Delta C_t$ method [55]. mRNA extractions were done using Trizol and reverse transcription steps were carried out using Fermentas kit. Finally, qPCR was performed using a Bio-Rad kit at the following step: 40 cycles of 94°C and 55°C. Primer sequences are available upon request.

events of molecular domestication, as suggested by the close phylogenetic relationship of *bona fide* GIN transposons with each of both genes (Figure 1). Presence of introns at conserved positions might in this case reflect intron conservation between ancestral GIN transposons at the origin of both molecular domestication events.

GIN1 and GIN2 functions might be related to the binding to DNA or RNA, since both proteins have conserved the HHCC zinc finger present in the ancestral integrase. Conservation of the integrase activity appears possible but must be tested through functional assays. In fish, *GIN2* is particularly expressed in brain and gonads; its expression pattern during zebrafish embryogenesis suggests a role during gastrulation. Functional analysis in fish will provide important insights into the biological function of *GIN2* in vertebrates.

Taken together, data on *GIN* and other TE-derived genes support the important role of molecular domestication as a driver of genetic innovation during evolution. What we have presented here probably only represents the tip of the

evolutionary iceberg. There is no doubt that future genome comparisons and functional gene analyses will uncover new domesticated genes and novel biological functions essential for the diversification of vertebrates and other living organisms.

Acknowledgments

The authors' work is supported by grants from the Agence Nationale de la Recherche (ANR).

References

- [1] W. F. Doolittle and C. Sapienza, "Selfish genes, the phenotype paradigm and genome evolution," *Nature*, vol. 284, no. 5757, pp. 601–603, 1980.
- [2] L. E. Orgel and F. H. C. Crick, "Selfish DNA: the ultimate parasite," *Nature*, vol. 284, no. 5757, pp. 604–607, 1980.
- [3] C. Feschotte and C. Gilbert, "Endogenous viruses: insights into viral evolution and impact on host biology," *Nature Reviews Genetics*, vol. 13, no. 4, pp. 283–296, 2012.

- [4] H. S. Malik, S. Henikoff, and T. H. Eickbush, "Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses," *Genome Research*, vol. 10, no. 9, pp. 1307–1318, 2000.
- [5] D. Ribet, F. Harper, A. Dupressoir, M. Dewannieux, G. Pierron, and T. Heidmann, "An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus," *Genome Research*, vol. 18, no. 4, pp. 597–609, 2008.
- [6] M. J. Curcio and K. M. Derbyshire, "The outs and ins of transposition: from MU to kangaroo," *Nature Reviews Molecular Cell Biology*, vol. 4, no. 11, pp. 865–877, 2003.
- [7] H. Kaessmann, "Origins, evolution, and phenotypic impact of new genes," *Genome Research*, vol. 20, no. 10, pp. 1313–1326, 2010.
- [8] J. N. Volff, "Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes," *BioEssays*, vol. 28, no. 9, pp. 913–922, 2006.
- [9] E. M. Zdobnov, M. Campillos, E. D. Harrington, D. Torrents, and P. Bork, "Protein coding potential of retroviruses and other transposable elements in vertebrate genomes," *Nucleic Acids Research*, vol. 33, no. 3, pp. 946–954, 2005.
- [10] M. Campillos, T. Doerks, P. K. Shah, and P. Bork, "Computational characterization of multiple Gag-like human proteins," *Trends in Genetics*, vol. 22, no. 11, pp. 585–589, 2006.
- [11] J. Brandt, S. Schrauth, A. M. Veith et al., "Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals," *Gene*, vol. 345, no. 1, pp. 101–111, 2005.
- [12] C. Charlier, K. Segers, L. Karim et al., "The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status," *Nature Genetics*, vol. 27, no. 4, pp. 367–369, 2001.
- [13] R. Ono, S. Kobayashi, H. Wagatsuma et al., "A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21," *Genomics*, vol. 73, no. 2, pp. 232–237, 2001.
- [14] S. Suzuki, R. Ono, T. Narita et al., "Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting," *PLoS Genetics*, vol. 3, no. 4, article e55, 2007.
- [15] R. Ono, K. Nakamura, K. Inoue et al., "Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality," *Nature Genetics*, vol. 38, no. 1, pp. 101–106, 2006.
- [16] Y. Sekita, H. Wagatsuma, K. Nakamura et al., "Role of retrotransposon-derived imprinted gene, Rtl1, in the fetomaternal interface of mouse placenta," *Nature Genetics*, vol. 40, no. 2, pp. 243–248, 2008.
- [17] M. Schüller, D. Jenne, and R. Voltz, "The human PNMA family: novel neuronal proteins implicated in paraneoplastic neurological disease," *Journal of Neuroimmunology*, vol. 169, no. 1–2, pp. 172–176, 2005.
- [18] J. Dalmau, S. H. Gultekin, R. Voltz et al., "Ma1, a novel neuron- and testis-specific protein, is recognized by the serum of patients with paraneoplastic neurological disorders," *Brain*, vol. 122, pp. 27–39, 1999.
- [19] H. L. Chen and S. R. D'Mello, "Induction of neuronal cell death by paraneoplastic Ma1 antigen," *Journal of Neuroscience Research*, vol. 88, no. 16, pp. 3508–3519, 2010.
- [20] K. O. Tan, N. Y. Fu, S. K. Sukumaran et al., "MAP-1 is a mitochondrial effector of Bax," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 41, pp. 14623–14628, 2005.
- [21] L. C. Edelstein and T. Collins, "The SCAN domain family of zinc finger transcription factors," *Gene*, vol. 359, no. 1–2, pp. 1–17, 2005.
- [22] R. O. Emerson and J. H. Thomas, "Gypsy and the birth of the SCAN domain," *Journal of Virology*, vol. 85, no. 22, pp. 12043–12052, 2011.
- [23] D. Ivanov, J. R. Stone, J. L. Maki, T. Collins, and G. Wagner, "Mammalian SCAN domain dimer is a domain-swapped homolog of the HIV capsid C-terminal domain," *Molecular Cell*, vol. 17, no. 1, pp. 137–143, 2005.
- [24] T. L. Sander, K. F. Stringer, J. L. Maki, P. Szauter, J. R. Stone, and T. Collins, "The SCAN domain defines a large family of zinc finger transcription factors," *Gene*, vol. 310, no. 1–2, pp. 29–38, 2003.
- [25] S. Best, P. L. Tissier, G. Towers, and J. P. Stoye, "Positional cloning of the mouse retrovirus restriction gene Fv1," *Nature*, vol. 382, no. 6594, pp. 826–829, 1996.
- [26] H. S. Malik, "Retroviruses push the envelope for mammalian placentation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 7, pp. 2184–2185, 2012.
- [27] M. Sha, X. Lee, X. P. Li et al., "Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis," *Nature*, vol. 403, no. 6771, pp. 785–789, 2000.
- [28] S. Blaise, N. De Parseval, L. Bénit, and T. Heidmann, "Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 22, pp. 13013–13018, 2003.
- [29] A. Dupressoir, G. Marceau, C. Vernochet et al., "Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 3, pp. 725–730, 2005.
- [30] O. Heidmann, C. Vernochet, A. Dupressoir, and T. Heidmann, "Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new "syncytin" in a third order of mammals," *Retrovirology*, vol. 6, article 107, 2009.
- [31] C. Vernochet, O. Heidmann, A. Dupressoir et al., "A syncytin-like endogenous retrovirus envelope gene of the guinea pig specifically expressed in the placenta junctional zone and conserved in Caviomorpha," *Placenta*, vol. 32, no. 11, pp. 885–892, 2011.
- [32] G. Cornelis, O. Heidmann, S. Bernard-Stoecklin et al., "Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 7, pp. E432–E441, 2012.
- [33] J. M. Antony, K. K. Ellestad, R. Hammond et al., "The human endogenous retrovirus envelope glycoprotein, syncytin-1, regulates neuroinflammation and its receptor expression in multiple sclerosis: a role for endoplasmic reticulum chaperones in astrocytes," *Journal of Immunology*, vol. 179, no. 2, pp. 1210–1224, 2007.
- [34] K. Søre, T. L. Andersen, A. S. Hobolt-Pedersen, B. Bjerregaard Bolette, L. I. Larsson, and J. M. Delaissé, "Involvement of human endogenous retroviral syncytin-1 in human osteoclast fusion," *Bone*, vol. 48, no. 4, pp. 837–846, 2011.

- [35] H. Ikeda and H. Sugimura, "Fv-4 resistance gene: a truncated endogenous murine leukemia virus with ecotropic interference properties," *Journal of Virology*, vol. 63, no. 12, pp. 5405–5412, 1989.
- [36] A. Marco and I. Marín, "CGIN1: a retroviral contribution to mammalian genomes," *Molecular Biology and Evolution*, vol. 26, no. 10, pp. 2167–2170, 2009.
- [37] T. Matsui, K. Miyamoto, A. Kubo et al., "SASPase regulates stratum corneum hydration through profilaggrin-to-filaggrin processing," *EMBO Molecular Medicine*, vol. 3, no. 6, pp. 320–333, 2011.
- [38] T. H. Eickbush, "Telomerase and retrotransposons: which came first?" *Science*, vol. 277, no. 5328, pp. 911–912, 1997.
- [39] A. Böhne, F. Brunet, D. Galiana-Arnoux, C. Schultheis, and J. N. Volff, "Transposable elements as drivers of genomic and biological diversity in vertebrates," *Chromosome Research*, vol. 16, no. 1, pp. 203–215, 2008.
- [40] C. Feschotte and E. J. Pritham, "DNA transposons and the evolution of eukaryotic genomes," *Annual Review of Genetics*, vol. 41, pp. 331–368, 2007.
- [41] V. V. Kapitonov and J. Jurka, "RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons," *PLoS Biology*, vol. 3, no. 6, article e181, 2005.
- [42] T. Okada, J. I. Ohzeki, M. Nakano et al., "CENP-B controls centromere formation depending on the chromatin context," *Cell*, vol. 131, no. 7, pp. 1287–1300, 2007.
- [43] C. Casola, D. Hucks, and C. Feschotte, "Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals," *Molecular Biology and Evolution*, vol. 25, no. 1, pp. 29–41, 2008.
- [44] H. P. Cam, K. I. Noma, H. Ebina, H. L. Levin, and S. I. S. Grewal, "Host genome surveillance for retrotransposons by transposon-derived proteins," *Nature*, vol. 451, no. 7177, pp. 431–436, 2008.
- [45] M. Zariatigui, M. W. Vaughn, D. V. Irvine et al., "CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR," *Nature*, vol. 469, no. 7328, pp. 112–115, 2011.
- [46] A. F. A. Smit and A. D. Riggs, "Tiggers and other DNA transposon fossils in the human genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 4, pp. 1443–1448, 1996.
- [47] W. Liu, J. Seto, G. Donovan, and M. Toth, "Jerky, a protein deficient in a mouse epilepsy model, is associated with translationally inactive mRNA in neurons," *Journal of Neuroscience*, vol. 22, no. 1, pp. 176–182, 2002.
- [48] F. Butter, D. Kappei, F. Buchholz, M. Vermeulen, and M. Mann, "A domesticated transposon mediates the effects of a single-nucleotide polymorphism responsible for enhanced muscle growth," *EMBO Reports*, vol. 11, no. 4, pp. 305–311, 2010.
- [49] E. Markljung, L. Jiang, J. D. Jaffe et al., "ZBED6, a novel transcription factor derived from a domesticated DNA transposon regulates IGF2 expression and muscle growth," *PLoS Biology*, vol. 7, no. 12, Article ID e1000256, 2009.
- [50] R. Cordaux, S. Udit, M. A. Batzer, and C. Feschotte, "Birth of a chimeric primate gene by capture of the transposase gene from a mobile element," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 21, pp. 8101–8106, 2006.
- [51] M. Shaheen, E. Williamson, J. Nickoloff, S. H. Lee, and R. Hromas, "Metnase/SETMAR: a domesticated primate transposase that enhances DNA repair, replication, and decatenation," *Genetica*, vol. 138, no. 5, pp. 559–566, 2010.
- [52] A. Sarkar, C. Sim, Y. S. Hong et al., "Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences," *Molecular Genetics and Genomics*, vol. 270, no. 2, pp. 173–180, 2003.
- [53] J. C. Newman, A. D. Bailey, H. Y. Fan, T. Pavelitz, and A. M. Weiner, "An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in cockayne syndrome," *PLoS Genetics*, vol. 4, no. 3, Article ID e1000031, 2008.
- [54] V. V. Kapitonov and J. Jurka, "Harbinger transposons and an ancient HARBI1 gene derived from a transposase," *DNA and Cell Biology*, vol. 23, no. 5, pp. 311–324, 2004.
- [55] A. Böhne, Q. Zhou, A. Darras et al., "Zisupton—a novel superfamily of DNA transposable elements recently active in fish," *Molecular Biology and Evolution*, vol. 29, no. 2, pp. 631–645, 2012.
- [56] S. E. Hammer, S. Strehl, and S. Hagemann, "Homologs of *Drosophila* P transposons were mobile in zebrafish but have been domesticated in a common ancestor of chicken and human," *Molecular Biology and Evolution*, vol. 22, no. 4, pp. 833–844, 2005.
- [57] T. Clouaire, M. Roussigne, V. Ecochard, C. Mathe, F. Amalric, and J. P. Girard, "The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 19, pp. 6907–6912, 2005.
- [58] J. B. Parker, S. Palchaudhuri, H. Yin, J. Wei, and D. Chakravarti, "A transcriptional regulatory role of the THAP11-HCF-1 complex in colon cancer cell function," *Molecular and Cellular Biology*, vol. 32, no. 9, pp. 1654–1670, 2012.
- [59] M. Roussigne, C. Cayrol, T. Clouaire, F. Amalric, and J. P. Girard, "THAP1 is a nuclear proapoptotic factor that links prostate-apoptosis-response-4 (Par-4) to PML nuclear bodies," *Oncogene*, vol. 22, no. 16, pp. 2432–2442, 2003.
- [60] M. Roussigne, S. Kossida, A. C. Lavigne et al., "The THAP domain: a novel protein motif with similarity to the DNA-binding domain of P element transposase," *Trends in Biochemical Sciences*, vol. 28, no. 2, pp. 66–69, 2003.
- [61] A. Sabogal, A. Y. Lyubimov, J. E. Corn, J. M. Berger, and D. C. Rio, "THAP proteins target specific DNA sites through bipartite recognition of adjacent major and minor grooves," *Nature Structural & Molecular Biology*, vol. 17, no. 1, pp. 117–123, 2010.
- [62] C. Lloréns and I. Marín, "A mammalian gene evolved from the integrase domain of an LTR retrotransposon," *Molecular Biology and Evolution*, vol. 18, no. 8, pp. 1597–1600, 2001.
- [63] I. Marín, "GIN transposons: genetic elements linking retrotransposons and genes," *Molecular Biology and Evolution*, vol. 27, no. 8, pp. 1903–1911, 2010.
- [64] W. Bao, V. V. Kapitonov, and J. Jurka, "Ginger DNA transposons in eukaryotes and their evolutionary relationships with long terminal repeat retrotransposons," *Mobile DNA*, vol. 1, no. 1, article 3, 2010.
- [65] S. Guindon, J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel, "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0," *Systematic Biology*, vol. 59, no. 3, pp. 307–321, 2010.
- [66] M. A. Larkin, G. Blackshields, N. P. Brown et al., "Clustal W and clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007.

- [67] H. S. Malik and T. H. Eickbush, "Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons," *Journal of Virology*, vol. 73, no. 6, pp. 5186–5190, 1999.
- [68] C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, and T. F. Schilling, "Stages of embryonic development of the zebrafish," *Developmental Dynamics*, vol. 203, no. 3, pp. 253–310, 1995.