Cell. Mol. Life Sci. 66 (2009) 1073 – 1093
1420-682X/09/061073-21
DOI 10.1007/s00018-009-8376-3
© Birkhäuser Verlag, Basel, 2009

Cellular and Molecular Life Sciences

Review

Molecular domestication of transposable elements: From detrimental parasites to useful host genes

L. Sinzelle[a,b], Z. Izsvák[a,c] and Z. Ivics[a,†,*]

[a] Max Delbrück Center for Molecular Medicine, 13092 Berlin (Germany)
[b] Epigenomics Program, Genopole®, CNRS, Université d'Evry Val d'Essonne, Evry (France)
[c] Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, 6726 Szeged (Hungary)

Received 2 July 2008; received after revision 1 October 2008; accepted 14 October 2008
Online First 12 January 2009

Abstract. Transposable elements (TEs) are commonly viewed as molecular parasites producing mainly neutral or deleterious effects in host genomes through their ability to move. However, during the past two decades, major interest has been focusing on the positive contribution of these elements in the evolution of gene regulation and in the creation of diverse structural host genes. Indeed, DNA transposons carry an attractive and elaborate enzymatic machinery as well as DNA components that have been co-opted in several cases by the host genome *via* an evolutionary process referred to as molecular domestication. A large number of transposon-derived genes known to date have been recruited by the host to function as transcriptional regulators; however, the biological role of the majority of them remains undetermined. Our knowledge on the structure, distribution, evolution and mechanism of transposons will continue to provide important contributions to our understanding of host genome functions.

Keywords. Transposon, diversity, evolution, transposase, DNA-binding domain, transcription factor, gene regulatory network.

Introduction

Transposable elements (TEs) are mobile, repetitive, genetic elements that are major components of all eukaryotic genomes investigated so far. The recent availability of complete eukaryotic genome sequences has considerably enriched the repertoire of annotated TEs, and revealed their abundance and great diversity. Two classes of transposon are distinguished according to their respective transposition mechanisms [1]. The mobility of class I elements or retrotransposons is achieved through an RNA intermediate mediating a "copy-and-paste" mechanism, and class II or DNA transposons use a DNA-mediated, "cut-and-paste" mode of transposition. Both classes exist as non-autonomous and autonomous elements. Autonomous copies encode all the enzymes necessary to move, whereas nonautonomous copies have no coding capacity, and therefore their mobility is entirely dependent on the enzymatic machinery of their autonomous relatives. TE-derived sequences make up about 45 % of the human genome, of which retrotransposons form the major type of TEs, whereas DNA transposons contribute to 3 % of the genome [2].

[†] Present address: Max Delbrück Center for Molecular Medicine, Robert-Rössle Strasse 10, 13092 Berlin (Germany), Fax: +49-30-94062547, e-mail: zivics@mdc-berlin.de
[*] Corresponding author.

Table 1. Structural and molecular properties of the nine superfamilies of DNA transposons belonging to the terminal inverted repeat (TIR) order.

Subclass 1 / Order TIR								
Superfamily	IS-related	Occurrence	Length (kb)	TIRs (bp)	TSDs (bp)	Encoded proteins	DBD	Catalytic core
Tc1/ mariner	*IS630*	Eukaryotes	1.2-5.0	17-1100	2 (TA)	Tnp	HTH	(D, D, E/D) Tnp
hAT	*nd*	Eukaryotes	2.5-5.0	5–27	8	Tnp	BED ZnF	(D, D, E) Tnp
Mutator	*IS256*	Eukaryotes	1.3-7.4	0-several kb	9–11	Tnp	WRKY/GCM1 ZnF	(D, D, E) Tnp
Merlin	*IS1016*	Animals and eubacteria	1.4-3.5	21-462	8–9	Tnp	nd	(D, D, E) Tnp
Transib	*nd*	Metazoans and fungi	3–4	9–60	5	Tnp	nd	(D, D, E) Tnp
P	*nd*	Plants and metazoans	3–11	13-150	8	Tnp	THAP ZnF	nd
piggyBac	*IS1380*	Eukaryotes	2.3-6.3	12–19	4 (TTAA)	Tnp	nd	nd
PIF/ Harbinger	*IS5*	Eukaryotes	2.3-5.5	15-270	3 (CWG or TWA)	Tnp + Myb-like protein	Myb/SANT	(D, D, E) Tnp
CACTA	*nd*	Plants, metazoans and fungi	4.5-15	10-54	2–3	TnpA + TnpD	nd	nd

To facilitate annotation of the growing data on TEs, a novel hierarchical classification system of eukaryotic class II elements based on transposition mechanism, sequence similarities and structural relationships has been recently proposed [3]. Class II elements are subdivided into two subclasses based on the generation of single- or double-stranded DNA cuts during the transposition process. Subclass 1 comprises cut-and-paste TEs that are flanked by terminal inverted repeats (TIRs) (Table 1). This TIR order is composed so far of nine superfamilies distinguished by the sequence motifs within their TIRs, and the length of the target site duplications (TSDs) resulting from the duplication of a short host DNA sequence generated flanking both transposon ends upon insertion. The two, recently identified *Helitron* and *Maverick* transposon families belong to a second subclass of DNA transposon, since their transposition process requires replication and does not introduce double-strand DNA breaks [4, 5]. Classical eukaryotic DNA transposons encode at least one enzyme, the transposase, that carries out the cut-and-paste transposition reaction *via* its two functional domains: an N-terminal DNA-binding domain (DBD) that recognizes and binds specifically to the transposon ends (TIRs and/or subterminal sequences) and a C-terminal catalytic domain that catalyzes both the DNA cleavage and strand transfer steps (reviewed in [6]) (Fig. 1A). The two superfamilies *CACTA* and *PIF/Harbinger* produce a second protein necessary for transposition [7–9]. Similar, the autonomous maize element *Mutator MuDR* contains two genes: *mudrA* encoding MURA transposase and *mudrB* whose product is required for transposon integration [10, 11]. However, the presence of *mudrB* is not a general feature within the *Mutator* family; it was found only in the genus *Zea* (reviewed in [12]). Except for the three superfamilies *P*, *piggyBac* and *CACTA* for which the catalytic domain is not yet well established, eukaryotic transposases carry a well-conserved [D, D, E/D] motif also found in retroviral integrases [13]. In addition, the amino acid spacing between the second D and the last D/E residues is specific for each superfamily. The [D, D, E/D] motif coordinates a metal ion that is specifically required for the nicking process and strand transfer reactions of the integration step (reviewed in [14]).

TEs are commonly viewed as selfish or parasitic entities, existing only to propagate themselves, independently of any beneficial effect on their host. The current model of their life-cycle consists of invading new species, increasing copy number, persisting within the genome until an ultimate phase in which elements exist as fossils [15]. Consistent with the selfish DNA theory, mobility of TEs produces a variety of detrimental effects, including insertional mutagenesis, leading to gene inactivation or expression pattern modification. In addition, the presence of several repeated sequences dispersed within the genome provides substrates for illegitimate recombination, creating chromosomal translocations, inversions, or deletions. However, our perception of the selfish nature of TEs has considerably evolved during the past two decades as a result of increasing numbers of studies that described the capacity of these elements as an important force in the evolution of gene regulation and in the creation of genetic novelty. Indeed, the literature describes several examples of TEs that donated promoters or enhancer sequences to host genes, as well as their contribution to provide alternative splice sites, polyadenylation sites and *cis*-regulatory sequences (reviewed in [16, 17]).

Another consequence of the intimate relationship between transposon and host genome is the creation of chimeric genes that can in some cases give rise to a functional protein. In *Drosophila*, one particular insertion of *P* element has been shown to produce a chimeric gene encoding the DBD of the *P* element and a functional domain of the target host gene [18]. Several genetic processes that lead to the formation of chimeric genes have been highlighted in plants. As an example, the alternative transposition of the maize *Ac/Ds* element from the *hAT* superfamily that involves the 5'- and 3'-ends of different elements has been shown to provoke the fusion of the coding sequence of two genes generating a functional chimeric gene and subsequently a new phenotype [19]. In rice, 3000 chimeric elements called Pack-MULEs that had captured >1000 gene fragments from different chromosomal loci have been detected [20]. However, the origins and the roles of these chimeric proteins remain enigmatic. Similarly, such transposon-induced rearrangements of large-scale duplication and shuffling of coding sequences have been reported for other *Mutator* elements, *Helitrons* and *CACTA* transposons (reviewed in [21]).

The great contribution of TEs on the evolution of a protein coding region was fully appreciated recently with large-scale *in silico* studies performed on the vast number of sequences available from model organisms, and from human [22, 23]. Indeed, it has been reported that TEs or TE fragments have contributed to at least 4% of human protein-coding genes [24, 25]. The majority of TEs were found to be distinct exons recruited into coding regions by splicing. Thus, it appears that in many instances, TEs and host genome have evolved a mutually beneficial relationship that balance TE survival and the evolutionary interest of the host.

The most striking beneficial contribution of TEs is illustrated by an evolutionary process referred to as “molecular domestication”, by which a TE-derived coding sequence gives rise to a functional host gene. Thus, domesticated genes represent stable functional components of the genome. Such transposon-derived genes were first identified as domesticated *P* elements in *Drosophila* [26] and further extended to plant and animal genomes, including human [27–29]. Preliminary sequence analysis of the human genome identified 47 TE-derived genes with a likely origin in up to 38 different transposon copies [2]. For instance, domesticated genes are known to have derived from almost all superfamilies of DNA transposons with the exception of *CACTA* and *Merlin* superfamilies. Several criteria have been proposed to determine strong cases of DNA transposon-derived genes [30]. In contrast to the repetitive nature of TEs, domesticated genes exist as single copies in the genome, and orthologs are detectable in distantly related species. Structurally, these genes are devoid of the molecular hallmarks of transposition such as flanking TIRs and TSDs. The protein products of domesticated genes are phylogenetically linked to transposon-encoded proteins. They assume important biological roles *in vivo* but, in general, they have lost their capacity to mediate transposition.

The aim of this review is to provide an overview of the vast repertoire of transposon-derived genes identified so far. This review focuses on domesticated transposases (or functional domains thereof) encoded by class II DNA transposons. First, we present representative strong cases of molecular domestication that illustrate the structural diversity of the emerging genes. The second part is devoted to the different evolutionary mechanisms that have led to the emergence of transposon-derived genes. Finally, the functional roles of these proteins involved in diverse biological processes (cell proliferation, apoptosis, cell cycle progression, chromosome segregation, chromatin modification, transcriptional regulation) are discussed with respect to the importance that transposons played in genome evolution and function.

Structural diversity

The increasing number of newly discovered domesticated genes clearly highlights their structural diversity. Some of these genes have emerged from the entire coding sequence of the transposase or exist as chimeric genes, in which the entire coding sequence of the transposase has been fused to a preexisting functional domain (Fig. 2). Furthermore, the structural diversity is reinforced by the fact that many domesticated genes have retained only the DBD domain of the ancestral transposon-encoded protein (Fig. 2). The following section describes some instances illustrating the great structural diversity of domesticated genes.

Molecular domestication of entire transposase genes

Several cases of host proteins derived from the complete coding sequence of the transposase have been reported, including human proteins such as the

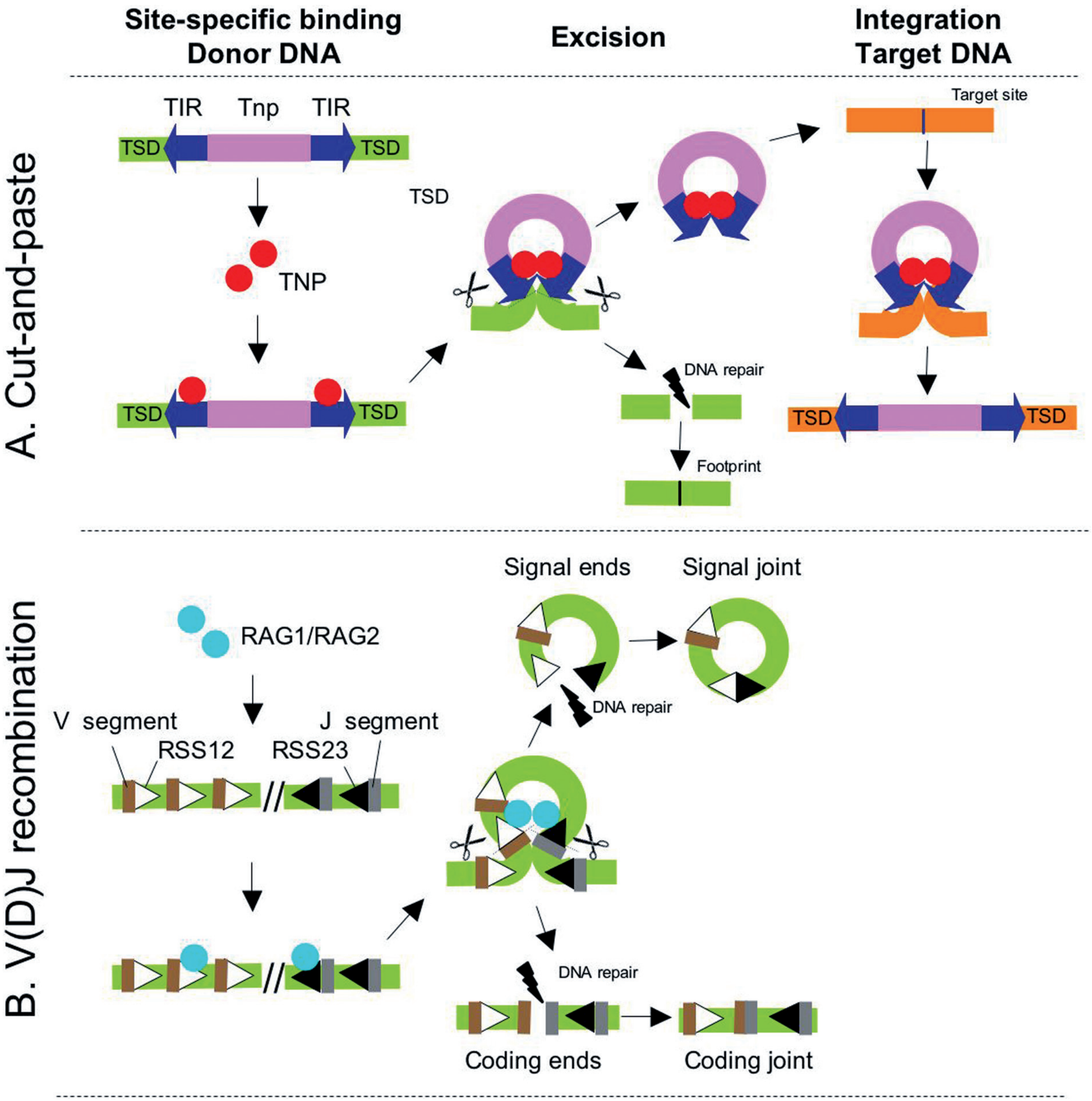

Figure 1. Functional homology between classical cut-and-paste transposition and V(D)J recombination. (*A*) Scheme of the classical cut-and-paste transposition process. An autonomous transposon consists of a coding region for the transposase (Tnp, pink rectangle), flanked on both ends by terminal inverted repeats (TIRs, blue arrows). The TIRs are flanked by target site duplications (TSDs), characteristic to each transposon family. The transposase protein (red sphere) specifically binds to its recognition sequences at each end of the transposon. The transposase excises the transposon by cleaving the DNA at the ends of the TIRs following formation of a synaptic complex. The cellular DNA repair machinery seals the excision site, and generates a transposon footprint of different length characteristic to each transposon family. The transposase recognizes a target site, and integrates the transposon into the target DNA, upon which the target site gets duplicated. (*B*) Schematic representation of a V(D)J recombination reaction. The brown and gray bars indicate V and J coding segments, respectively. Each J segment is associated with an RSS23 (black triangles) and each V segment with an RRS12 (open triangles). Recombination initially requires specific binding of the RAG1/RAG2 recombinase to a 12/23 RSS pair. RAG1/RAG2 form a synaptic complex, in which the two DNA strands immediately adjacent to each RSS are cleaved and processed by a nick-hairpin mechanism. The double-stranded breaks in the coding DNA are repaired to give rise to coding joints. Signal ends are joined together to generate signal joints, which are lost from the cell.

centromeric protein B, *i.e.*, CENP-B, Jerky and ZBED1 (zinc finger BED domain containing protein 1) proteins (Table 2). Here, we focus on two examples of well-characterized, transposon-derived proteins, the recombination-activating gene products RAG1/2 in vertebrates and the two homologous genes far-red impaired response protein 1 (FAR1) and far-red elongated hypocotyl 3 (FHY3) in plants.

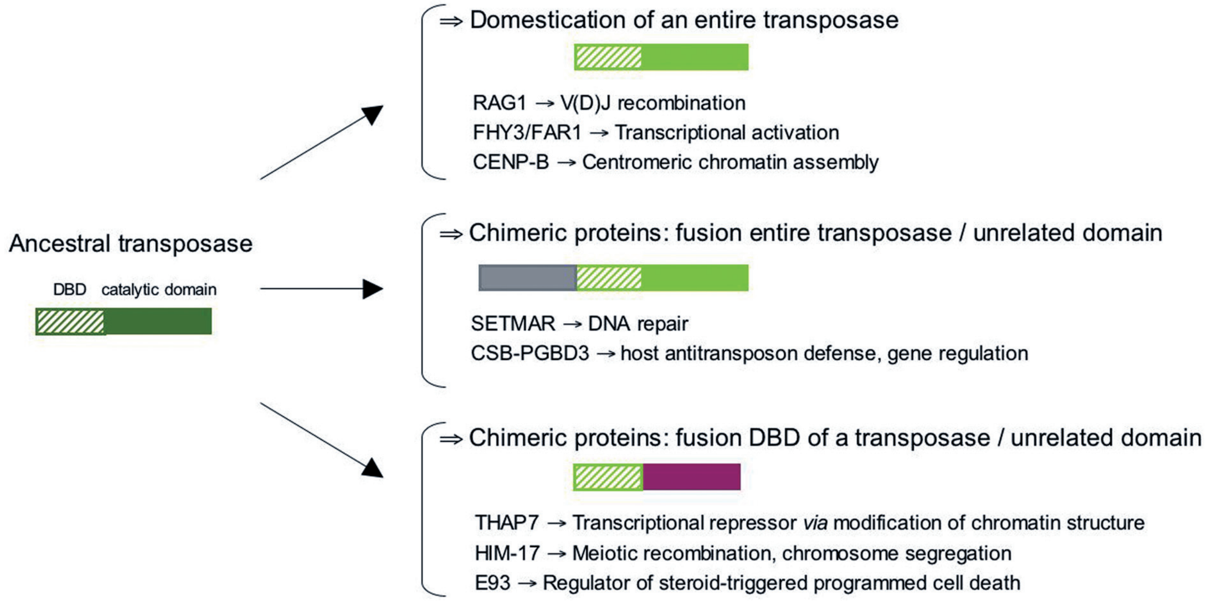
Ancestral transposase
DBD catalytic domain
⇒ Domestication of an entire transposase
RAG1 → V(D)J recombination
FHY3/FAR1 → Transcriptional activation
CENP-B → Centromeric chromatin assembly
⇒ Chimeric proteins: fusion entire transposase / unrelated domain
SETMAR → DNA repair
CSB-PGBD3 → host antitransposon defense, gene regulation
⇒ Chimeric proteins: fusion DBD of a transposase / unrelated domain
THAP7 → Transcriptional repressor *via* modification of chromatin structure
HIM-17 → Meiotic recombination, chromosome segregation
E93 → Regulator of steroid-triggered programmed cell death

Figure 2. Structural diversity of domesticated proteins. Classical transposase proteins contain a DNA-binding domain (DBD) (hatched green rectangle) and a catalytic domain (green rectangle). Domestication events of a transposase can give rise to diverse structural proteins: domestication of an entire transposase gene, chimeric genes formed by an entire transposase domain and an additional functional domain, and chimeric genes formed by the DBD of a transposase and an additional functional domain. For each of these three cases, some domesticated proteins and their respective functional role(s) are provided as examples.

Vertebrates

V(D)J recombination, a site-specific recombination reaction in the immune system of jawed vertebrates is incontestably the most spectacular example that TEs can derive complex and crucial functions in the host. In this process, which occurs during lymphocyte development, preexisting V (variable), D (diversity), and J (joining) gene segments are rearranged to generate a large repertoire of T cell surface receptor (TCR) and immunoglobulin molecules necessary for the recognition of diverse pathogens. The recombination event involves *cis*-acting sequences known as recombination signal sequences (RSSs) that flank each receptor gene segment and two proteins encoded by the recombination-activating genes RAG1 and RAG2. RSSs consist of unique conserved heptamer and nonamer sequences separated by either 12 or 23 nucleotides (Fig. 1B). The site specificity of the recombination is defined by the binding of RAG1 to the RSS. Typically, the V(D)J recombination reaction is subdivided into two stages, a cleavage phase and a joining phase (reviewed in [31]). The complex formed by the RAG1 and RAG2 proteins introduces double-strand breaks in the DNA between the heptamer of the RSS and the neighboring coding DNA *via* a nick-hairpin mechanism. The reaction results in the formation of two hairpins at the coding end and two blunt signal ends by a transesterification mechanism. After opening of the hairpins, repair factors of the non-homologous end joining (NHEJ) pathway join the two coding DNA segments together to generate the mature receptor gene (coding joints), as well as the signal ends (signal joints) which are lost from the cell. Mechanistically, the V(D)J recombination reaction shares significant similarities with the excision step of the cut-and-paste transposition process by which the transposon is excised from the donor-site DNA *via* double-strand breaks [32]. Moreover, V(D)J recombination produces a hairpin intermediate formed at the ends of the broken donor DNA similar to that described in *Hermes* transposition [33]. *In vitro*, purified RAG proteins have the capacity to transpose a piece of DNA flanked by two RSSs into a target DNA [32, 34]. In addition, RAG transposition events can occur at low frequencies in yeast and mammalian cells [35–37]. RAG-mediated transposition predominantly produces 5-bp TSDs upon insertion (reviewed in [38]).

The link between DNA transposition and V(D)J recombination has also been emphasized with the analysis of the structural features of the V(D)J recombination components [38]. The C-terminal domain of RAG1 including the [D, D, E] catalytic triad, the structure of the RSSs as well as the characteristic TSDs strongly support that RAG1 and the RSSs originate from a formerly active *Transib* transposon. Recently, a novel transposon called *N-RAG-TP* identified from the sea slug *Aplysia california* was found to encode a protein similar to the N-terminal part of RAG1 in vertebrates, which further

Table 2. Structural features of domesticated proteins and their respective functional role(s) in various biological processes[a].

DNA transposon superfamily	Gene ID	Name	Original organism/ Distribution	DBD family	Additional domains or parts of genes	Functions	References
Tc1/ mariner Pogo	Abp1	ARS-binding protein 1	*Saccharomyces pombe*	HTH_CENP-B	–	chromosome segregation, centromeric heterochromatin formation, retrotransposition control	[124, 125]
	Bab1	bric à brac 1	*Drosophila melanogaster*	HTH_PSQ	BTB	development and morphogenesis of ovaries, legs, antennae and abdomen	[144]
	Bab2	bric à brac 2	*Drosophila melanogaster*	HTH_PSQ	BTB	development and morphogenesis of ovaries, legs, antennae and abdomen	[144]
	Cbh1	CENP-B homolog 1	*Saccharomyces pombe*	HTH_CENP-B	–	chromosome segregation, centromeric heterochromatin formation, retrotransposition control	[123, 125]
	Cbh2	CENP-B homolog 2	*Saccharomyces pombe*	HTH_CENP-B	–	chromosome segregation and centromeric heterochromatin formation	[123, 125]
	CENP-B	Centromere protein B	*Homo sapiens /* mammals	HTH_CENP-B	–	centromeric chromatin assembly	[116]
	Eip93F	Drosophila cell death protein E93	*Drosophila melanogaster*	HTH_PSQ	–	regulator of steroid-triggered programmed cell death during metamorphosis	[88]
	JRK	Jerky	*Homo sapiens /* mammals	HTH_CENP-B	–	DNA-and RNA-binding activity in neurons	[93]
	JRKL	Jerky-like	*Homo sapiens/* mammals	HTH_CENP-B	–	unknown	[93]
	Pdc2	pyruvate decarboxylase 2	*Saccharomyces cerevisiae / Saccharomycetales*	HTH_CENP-B	–	regulator of gene expression in pyruvate decarboxylase and thiamin metabolism	[145]
	Psq	pipsqueak	*Drosophila melanogaster*	HTH_PSQ	BTB	transcriptional repressor during embryonic and adult development	[130]
	SETMAR (=Metnase)	SET domain and *mariner* transposase fusion	*Homo sapiens /* Anthropoid/ Primates	HTH	SET	histone methyltransferase function, enhance resistance to ionizing radiation, DNA repair	[51-56]
Transib	RAG1	recombination-activating gene 1	*Homo sapiens*/ jawed vertebrates	nd	Zn_RING, NBR	V(D)J recombination	[38]
Mutator	Aft1	activator of ferrous transport 1	*Saccharomyces cerevisiae / Saccharomycetales*	Zn_WRKY/ GCM1	–	ion utilization and homeostasis	[129]
	FAR1	far-red impaired response protein 1	*Arabidopsis thaliana* /Eudicots	Zn_WRKY/ GCM1		transcriptional activator in Phytochrome A signalling pathway for far-red light sensing	[40, 41]
	FHY3	far-red elongated hypocotyl 3	*Arabidopsis thaliana* /Eudicots	Zn_WRKY/ GCM1		transcriptional activator in Phytochrome A signalling pathway for far-red light sensing	[40, 41]
	MUG1	Mustang1	*Arabidopsis thaliana / Angiospermes*	Zn_WRKY/ GCM1	PB1	unknown	[47]
	Rcs1		*Saccharomyces cerevisiae / Saccharomycetales*	Zn_WRKY/ GCM1	–	ion utilization and homeostasis	[146]

Table 2 *(Continued)*

DNA transposon superfamily	Gene ID	Name	Original organism/ Distribution	DBD family	Additional domains or parts of genes	Functions	References
	Rbf1	RPG-box-binding factor 1	*Saccharomyces cerevisiae / Saccharomycetales*	Zn_WRKY/ GCM1	–	ion utilization and homeostasis	[147]
hAT	BEAF-32	boundary element-associated factor of 32 kDa	*Drosophila melanogaster/ Drosophilidae*	Zn_BED	–	insulator activity, chromatin structure, gene regulation	[70, 76]
	DAYSLEEPER	DAYSLEEPER	*Arabidopsis thaliana*	Zn_BED	–	essential for plant development	[48]
	DREF	DNA replication-related element-binding factor	*Drosophila melanogaster/ Drosophilidae*	Zn_BED	–	DNA replication, cell proliferation, growth and differentiation	[77]
	Gary	Gary	grasses	nd	–	unknown	[49]
	GON-14	gonadogenesis deficient, lin15b family member	*Caenorhabditis elegans*	Zn_THAP	–	pleiotropic regulator of animal development	[67]
	GTF2IRD2	GTF2I repeat domain containing 2, fusion with GTF2I domain of TFII-I transcription factor	*Homo sapiens /* mammals	Zn_BED	GTF2I gene	may play a role in Williams-Beuren syndrome	[60]
	LIN-15B	abnormal cell LINeage family member 15B	*Caenorhabditis elegans*	Zn_BED	Zn_THAP	vulval development, cell proliferation, cell cycle G1/S inhibitor	[136]
	ZBED1 (=hDREF=TRAMP)	zinc finger BED domain containing protein 1, human homolog of DREF	*Homo sapiens /* vertebrates	Zn_BED	–	transcription factor, cell proliferation, regulation of ribosomal protein	[78, 80]
	ZBED4	zinc finger BED domain containing protein 4	*Homo sapiens /* vertebrates	Zn_BED	–	unknown	[28]
	ZBED5 (=Buster1)	zinc finger BED domain containing protein 5, fusion with part of eIF4G2 protein	*Homo sapiens /* mammals	Zn_BED	–	translational repressor, modulator of interferon-gamma-induced appotosis	[28]
P	CDC14B	cell-cycle regulator tyrosine phosphatase, isoform B	*Caenorhabditis elegans*	Zn_THAP	CDC14	cell cycle control, G1/S inhibitor, genome stability	[148]
	CTB-1	homolog of CtBP transcriptional corepressor	*Caenorhabditis elegans*	Zn_THAP	NAD_b	transcriptional corepressor for development and oncogenesis	[149]
	HIM-17	high incidence of males 7	*Caenorhabditis elegans*	Zn_THAP	coiled coil	chromatin modification, meiotic chromosome segregation	[131]
	LIN-36	abnormal cell LINeage family member 36	*Caenorhabditis elegans*	Zn_THAP	Zn_C2H2	vulval development, cell proliferation, cell cycle G1/S transition inhibitor	[136, 150]
	P-neo G and A type	obscura P-neogene	*Drosophila subobscura/ subobscura* subgroup	Zn_THAP	–	unknown	[26]
	P-neo montium	montium P-neogene	*Drosophila montium/montium* subgroup	Zn_THAP	–	unknown	[110, 111]
	P-boc	bocqueti stationary P-neogene	*Drosophila bocqueti*	Zn_THAP	–	unknown	[107]
	P-tsa	tsacasi stationary P-neogene	*Drosophila tsacasi*	Zn_THAP	–	unknown	[111]

Table 2 *(Continued)*

DNA transposon superfamily	Gene ID	Name	Original organism/ Distribution	DBD family	Additional domains or parts of genes	Functions	References
	THAP0(=DAP4=p52riPK)	thanatos associated protein 0, death-associated protein interferon-induced protein kinase-interacting protein	*Homo sapiens* / vertebrates	Zn_THAP	hATC	interferon-gamma-induced apoptosis	[132, 133]
	THAP1	nuclear proapoptotic factor THAP1	*Homo sapiens* / vertebrates	Zn_THAP	–	serum withdrawal- interferon-gamma-induced apoptosis	[134]
	THAP2	thanatos-associated protein 2	*Homo sapiens* / mammals-vertebrates	Zn_THAP	–	unknown	[63]
	THAP3	thanatos-associated protein 3	*Homo sapiens* / mammals-vertebrates	Zn_THAP	–	unknown	[63]
	THAP4	thanatos-associated protein 4	*Homo sapiens* / mammals-vertebrates	Zn_THAP	–	unknown	[63]
	THAP5	thanatos-associated protein 5	*Homo sapiens* / mammals-vertebrates	Zn_THAP	–	unknown	[63]
	THAP6	thanatos-associated protein 6	*Homo sapiens* / mammals-vertebrates	Zn_THAP	–	unknown	[63]
	THAP7	thanatos-associated protein 7	*Homo sapiens* / vertebrates	Zn_THAP	–	binds hypoacethylated histone H4 tails, recruits histone deacetylase HDAC3 and NcoR to specific DNA sites	[66]
	THAP8	thanatos-associated protein 8	*Homo sapiens* / mammals-vertebrates	Zn_THAP	–	unknown	[63]
	THAP9 (=Phsa)	thanatos-associated protein 9, P element-homologous gene	*Homo sapiens* / mammals-amniotes	Zn_THAP	–	unknown	[65, 112]
	THAP10	thanatos-associated protein 10	*Homo sapiens* / mammals-vertebrates	Zn_THAP	–	unknown	[63]
	THAP11	thanatos-associated protein 11	*Homo sapiens* / mammals-vertebrates	Zn_THAP	–	unknown	[63]
	THAP-E2F6	fusion of THAP and cell cycle transcription factor E2F6	*Danio rerio* /fishes, amphibians	Zn_THAP	E2F_TDP gene	repressor of E2F-dependent transcription during S phase	[137]
PIF/ Harbinger	DPLG1-7	Drosophila PIF-like genes 1-7	*Drosophila*	nd		unknown	[101]
	DPMG 7	Drosophila PIF MADF-like protein encoding gene 7	*Drosophila*	Myb/SANT/ trihelix		unknown	[101]
	HARBI1	Harbinger derived-protein 1	*Homo sapiens* / vertebrates	nd		interacts with NAIF1	[9, 100]

Table 2 *(Continued)*

DNA transposon superfamily	Gene ID	Name	Original organism/ Distribution	DBD family	Additional domains or parts of genes	Functions	References
	NAIF1	Nuclear apoptosis-inducing factor 1	*Homo sapiens* / vertebrates	Myb/SANT/ trihelix		interacts with and promotes nuclear import of HARBI1, and induces apoptosis when overexpressed	[9, 102]
CACTA	ROSINA	RSI	*Anthirinium majus*	nd	–	modulator of petal and stamen development	[151]
piggyBac	KOBUTA	KOBUTA	*Xenopus tropicalis/ laevis/borealis*	nd	–	unknown	[139]
	PGBD1(=HUCEP-4)	piggyBac-derived 1, cerebral protein 4	*Homo sapiens* / mammals-primates	nd	Zn_SCAN	unknown	[57]
	PGBD2	piggyBac-derived 2	*Homo sapiens* / mammals-primates	nd	–	unknown	[57]
	PGBD3	piggyBac-derived 3, Cockayne Syndrome group B gene and piggyBac transposase fusion	*Homo sapiens* / mammals-primates	nd	CSG gene	may play a role in Cockayne Syndrome	[57, 59]
	PGBD4	piggyBac-derived 4	*Homo sapiens* / mammals-primates	nd	–	unknown	[57]
	PGBD5	piggyBac-derived 5	*Homo sapiens* / mammals-primates	nd	–	unknown	[57]

[a] Domesticated proteins that have retained the DNA-binding domain (DBD) as well as the catalytic domain (partially or fully) of the ancestral transposase are shown in gray background. DBDs in alphabetical order: BED (BEAF and DREF); CENP-B (Centromere-binding protein B); HTH (helix-turn-helix); PSQ (pipsqueak); THAP (Thanatos-associated protein); WRKY/GCM1 (glial cell missing 1); Zn (zinc finger); nd (not determined). The additional genes or functional domains in alphabetical order: BTB (broad-complex, tramtrack, bric à brac); CSG (Cockayne syndrome group B); hATC (hAT C-terminal dimerization); NAD_b (NAD_binding); PB1 (Phox and BEM1); SET (suppressor of variegation, enhancer of zeste and trithorax).

supports the emergence of the V(D)J recombination machinery from transposons [39].

Plants

In plants, the homologous genes *FAR1*, *FHY3* and *FAR1–related sequence* (*FRS*) are transcription factors that modulate the phytochrome A (phyA) signaling pathway by activating transcription of *FHY1* and *FHL* whose products are essential for light-induced phyA nuclear accumulation and light response [40, 41]. These proteins contain three domains similar to those described in the *Mutator* transposase: an N-terminal C2H2-type zinc finger (ZF) motif of the WRKY-GCM1 family, a putative [D, D, E/D] central catalytic core and a C-terminal SWIM motif [42]. Evolutionary analysis of these genes has confirmed that the *FHY3/FAR1* gene family has been co-opted from one or several related *Mutator* elements [41]. The maize *Mutator* transposase MURA is known to regulate the expression of both its own genes and occasionally adjacent genes [43–45]. Thus, it appears that FHY3 and FAR1 have retained the transcriptional activity of an ancestral *Mutator* transposase. Interestingly, the JITA transposase encoded by the *Mutator* element *Jittery*, which is the closest homolog of FAR1 in the current databases, is active in excision but inactive in integration [40, 46]. In addition to *Far1/Fhy3*, several genes derived from complete transposase sequences have been identified in plants: *Mustang* that arose from the *Mutator* superfamily and *Daysleeper* and *Gary* that have emerged from the *hAT* superfamily [47–49]. No function has yet been ascribed to these proteins, although they are speculated to act as transcriptional regulators.

Chimeric genes emerged from fusions of entire transposase genes and additional functional domains

The primate-specific SETMAR [50] is a chimeric protein created by the fusion of a SET domain and the entire transposase-coding region of a *mariner*-like *Hsmar1* transposon [51]. SETMAR was shown to exhibit two biochemical functions *in vitro*, a histone methyltransferase function conferred by the SET domain and a DNA cleavage activity provided by the transposase domain [52–54]. In addition, the protein has retained many of the specific activities required for transposition including *Hsmar1* TIR-specific DNA-binding, formation of a paired-end complex, 5'-cleavage at the TIR and integration of precleaved DNA substrates into a TA dinucleotide target site [51, 53, 54]. However, SETMAR is defective in transposition due to its inability to achieve 3'-cleavage of the transposon ends, thereby generating only single-stranded nicks [53, 54]. Recently, it has been shown that the DNA-nicking activity of SETMAR is independent of its TIR-specific DNA binding [55].

Although SETMAR (together with its cellular interactor Pso4) has been suggested to play a role in DNA repair [56], and was shown to enhance resistance to ionizing radiation, its cellular functions remain poorly understood in either DNA repair or gene regulation by epigenetic modification [52]. Given the fact that *Hsmar1*-type transposons may provide ~7000 potential binding sites dispersed throughout the human genome, and that the biological functions of SETMAR are likely linked to its DNA-binding capacity, this protein has the potential to act as a regulator of gene expression controlling a vast network [51, 55].

Another instance of a chimeric domesticated gene is provided by the recent finding of five domesticated genes, *PGBD1–5* (*piggyBac-derived 1–5)*, derived from a *piggyBac* transposase in the human genome [57]. *PGBD1*, also referred to as cerebral protein 4 (HUCEP-4), is a chimeric protein formed by a C-terminal region derived from a *piggyBac* transposase and a SCAN-like domain, a highly conserved protein-protein interaction motif found near the N terminus of a subfamily of Cys2His2 ZF proteins (reviewed in [58]). *PGBD2–5* exhibit diverse structural organization [57]: *PGBD3* that comprises at least four pseudogenes as well as *PGBD2* are related to the *piggyBac* domain of *PGBD1,* and display two and one intron, respectively; *PGBD4* consists of a single ORF without introns, and is associated with the abundant nonautonomous *MER75* and *MER75B* transposons [27]. *PGBD5* is the most divergent element with the presence of eight introns. Recently, Newman et al. [59] have found that *PGBD3* may play a role in Cockayne syndrome. This domesticated gene, located in intron 5 of the Cockayne syndrome Group B (CSB) gene acts as an alternative 3'-terminal exon to produce a CSB-PGBD3 fusion protein. In addition, this fusion protein and the 3'-splice site are perfectly conserved in primate lineages. The authors speculated that the CSB-PGBD3 fusion protein together with the abundant nonautonoumous *MER83* elements may provide a gene regulatory network, similar to that proposed for the SETMAR protein and its putative genomic binding sites derived from the *Hsmar1* TIRs [51, 55].

GTF2IRD2 represents another example of a chimeric protein that has emerged from the fusion between a *Charlie8* transposase-like domain (*hAT* superfamily) and the GTF2I domain of the TFII-I transcription factor [60]. This fusion protein could be involved in the pathology of Williams-Beuren Syndrome.

Emergence of chimeric genes by recruitment of transposase DNA-binding domains

Recruitment of the DBD of a transposon-encoded protein appears to be a recurrent theme in the domestication of DNA transposons. Indeed, transposases are distinctive in possessing diverse DBDs such as ZF and helix-turn-helix (HTH) domains, providing a rich source for co-option by the host to give rise to chimeric genes that mainly act as transcription factors (Table 2).

Zinc-finger motifs

The THAP family (P transposase superfamily)

The most prominent example of DBD recruitment is the THAP (Thanatos-associated protein) domain, recently identified as a novel protein motif that harbors significant similarities with the site-specific DBD of the *Drosophila* canonical *P* protein, including its size of ~90 amino acid residues and its N-terminal position in the proteins [61, 62]. The THAP family is evolutionarily conserved from *Drosophila* to human, and comprises at least 12 members in humans (THAP0–11) as well as more than 100 distinct members in model animal organisms [61, 63]. This family includes the zebrafish orthologue of the cell-cycle regulator E2F6 and five *Caenorhabditis elegans* proteins, LIN-36, LIN-15A, LIN-15B, HIM-17 and GON-14. This domain is defined by well-conserved sequence motifs including an atypical ZF motif characterized by a C2CH module (consensus Cys-Xaa_{2-4}-Cys-Xaa_{35-50}-Cys-Xaa_2-His with a spacing of up to 53 residues between the zinc-coordinating C2 and CH residues) as well as a C-terminal AVPTIF box shown to be responsible for the site-specific DNA-binding activity of the *P* transposase [61, 64]. The evolutionary relationships between the THAP proteins and the *P* transposase were further supported by the significant sequence similarity between the human THAP9 protein (also referred to as *Phsa* [65]) and the *P* transposase through their entire sequences. THAP proteins have important roles in cell proliferation, cell-cycle control and apoptosis [63, 66, 67] and the THAP domain has been characterized as a ZF-based, sequence-specific DBD involved in transcriptional regulatory functions [63, 66, 68].

The BED domain (hAT transposase)

The BED (BEAF and DREF) domain was defined by Aravind [69] as a distinct DBD characterized by a Cys-Xaa_2-Cys-Xaa_n-His-Xaa_{3-5}-(H/C) signature in which X_n represents a variable spacer that is predicted to form a ZF. This domain is shared by BEAF-32 (boundary element-associated factor of 32 kDa [70]) and DREF (DRE-binding factor [71]) as well as transposases of the *hAT* superfamily. Based on sequence analysis, it has been proposed that the BED finger arose from transposases at two or more independent domestication events [69]. Indeed, it has been found that both BEAF and DREF possess DNA-binding activity [71, 72]. BEAF-32 binds to scs' insulator elements and to several hundred sites on polytene chromosomes in *Drosophila* [73, 74]. This binding is required for the insulator activity of the BEAF proteins that function by modulating chromatin structure [75, 76]. DREF was first described in *Drosophila* as a transcriptional regulator acting *via* specific binding at DRE (DNA replication-related element) sequences located in the promoters of many genes involved in DNA replication, cell growth and differentiation [71, 77]. The human ortholog of DREF called hDREF/KIAA0785 [78] (also called TRAMP [79] and ZBED1 [80]) is a transcription factor that binds to hDRE-like sequences, and regulates a set of human ribosomal protein genes [80]. hDRE-like sequences are also present in promoters of genes involved in cell proliferation and cell cycle progression, similar to that observed for DRE in *Drosophila.* DREF factors and *hAT* transposases share a C-terminal hATC (*hAT* C-terminal dimerization) domain that has been found to be a dimerization domain of *Activator* and *Hermes* transposases [81, 82]. Similarly, the hATC domain is necessary for hDREF self-association *in vivo*, and is also required for nuclear accumulation, DNA-binding activity and granular pattern formation [83].

The WRKY/GCM1 domain (Mutator transposase)

The WRKY/GCM1 superfamily of DBDs is also a striking example to illustrate the significant role of DNA transposons in the emergence of new transcription factors [42]. This superfamily of ZF proteins includes three major families of DBDs, namely WRKY, the DBD of the Glial Cell Missing (GCM1) transcription factors and FLYWCH, with DBDs of two distinct families of *Mutator* transposase. The transcription factors FAR1/FHY3 belong to this superfamily of ZF DBDs [42].

Helix-turn-helix motifs

The paired domain (Tc1 transposase)

The paired domain that characterizes the paired box (PAX) proteins is composed of two HTH subdomains: an N-terminal subdomain called PAI and a C-terminal subdomain called RED (PAI+RED=PAIRED). The early evolution of the paired domain has been reinvestigated by Breitling and Gerber [84], who proposed that the paired domain was originally derived from the DBD of an ancestral Tc1-like transposase.

The Pipsqueak family (Pogo transposase)

Another example of a DBD shared between cellular proteins and a transposase is Pipsqueak (Psq), a family of HTH proteins in eukaryotes that includes proteins from fungi, sea urchins, nematodes, insects and vertebrates [85]. This domain consists of four tandem repeats of a 50-amino acid sequence, in which each repeat represents a Psq motif. Within this family, three groups of proteins have been distinguished based on structural features and phylogenetic relationships [85]: (1) the BTB group that comprises proteins containing a protein-protein interaction domain called BTB (Broad-Complex, Tramtrack, Bric a brac)/POZ [86] and that includes the *Drosophila* Pipsqueak protein [87]; (2) the E93 group that contains the cell death regulator E93, a key regulator of steroid-triggered programmed cell death during *Drosophila* metamorphosis [88] as well as E93 orthologs found in coelenterates, nematodes and humans; and (3) the CENP-B/transposase group of two human proteins, the centromere-associated protein B (CENP-B) involved in centromeric heterochromatin assembly and the predicted protein CAB66474, as well as the *Drosophila Pogo* transposase belonging to the Tc1*/mariner* superfamily. It has been shown that the Psq motif of the *Pogo* transposase is responsible for the specific binding of transposon ends [89]. Nine other *Pogo*-derived genes have been identified and restricted to mammals: the *Tigger*-derived genes 1–7 (*TIGD1–7*) [90, 91], *Jerky* (*JRK*) that has both DNA and RNA-binding activity and is localized specifically in neurons [92, 93] and *Jerky-like* [94].

The Myb/SANT/trihelix domain (PIF/Harbinger Myb-like protein)

The Myb DBD was first described in the transcriptional regulator c-Myb involved in the control of cell proliferation and differentiation [95]. This domain consists of three imperfect tandem repeats (R1, R2 and R3), each containing three helices and characterized by regularly spaced tryptophan residues [96]. The SANT domain, identified based on its similarity with Myb-repeats, is found in many chromatin regulatory proteins [97] and is functionally involved in histone acetylation, deacetylation and ATP remodeling (reviewed in [98]). However, no DNA-binding activity has been reported for the SANT domain. The autonomous *PIF/Harbinger* transposons are known to encode a transposase and a second protein (referred to as the Myb-like protein) that contains a Myb/SANT/trihelix motif [99, 100, 101]. It was recently shown that the Myb/SANT/trihelix motif of the Myb-like protein functions as a DBD that specifically recognizes binding sites in both ends of the *Harbinger3_DR* transposon [9]. It was also found that the Myb/SANT/trihelix motif of a Myb-like protein has been domesticated ~500 million years ago in a common ancestor of jawed vertebrates to give rise to NAIF1 (nuclear apoptosis-inducing factor 1) [9], a pro-apoptotic protein [102].

Evolutionary diversity in domestication events

Before we discuss the possible evolutionary events that led to the emergence of transposon-derived host genes, it has to be mentioned that evolution often works by convergence and, therefore, it is at least formally possible that transposases could have evolved from host genes, not only the other way around. There are two criteria that can be used as a general guide when assessing host gene–transposon evolutionary relationships. The first is sequence similarity. In general, a high (>15 %) identity between relatively long (>200 amino acids) proteins is strong evidence for the transposon→host gene pathway. However, the common catalytic centers in evolutionarily unrelated proteins could have arisen as a result of convergent evolution. For instance, the [D, D, E/D] catalytic triad in Tc1/*mariner*, *MuDR*, *Harbinger*, *hAT*, and *Transib* transposases might have evolved convergently in each superfamily (there is no significant protein identity between transposases from different superfamilies). The second criterion is position in the phylogenetic tree. For example, if transposase-derived host genes are of relatively recent origin, but are related to a group of transposases with a much broader phylogenetic distribution, then it can be inferred that the host gene is derived from the transposase rather than *vice versa.* However, if the host gene has a deeper phylogenetic origin, it may be more difficult to infer evolutionary relationships to transposases. For example, the PAIRED domain found in Tc1 transposases and PAX proteins is of ancient origin, but the transposases seem to have a deeper evolutionary origin. Thus, in this case it seems more likely that the PAIRED domain was derived from a transposase and not the other way around (see ’Helix-turn-helix motifs’ above).

Multiple acquisition events

P element domestication is a unique example of multiple independent acquisitions of the same TE-derived coding sequence that had occurred in separate lineages of *Drosophila* (Fig. 3A; reviewed in [103]). Two distinct classes of functional *P* elements have been distinguished [104]. The first class represents canonical *P* DNA transposons of drosophilid flies. The transposase gene consisting of four exons is regulated in a tissue-specific manner by alternative splicing of

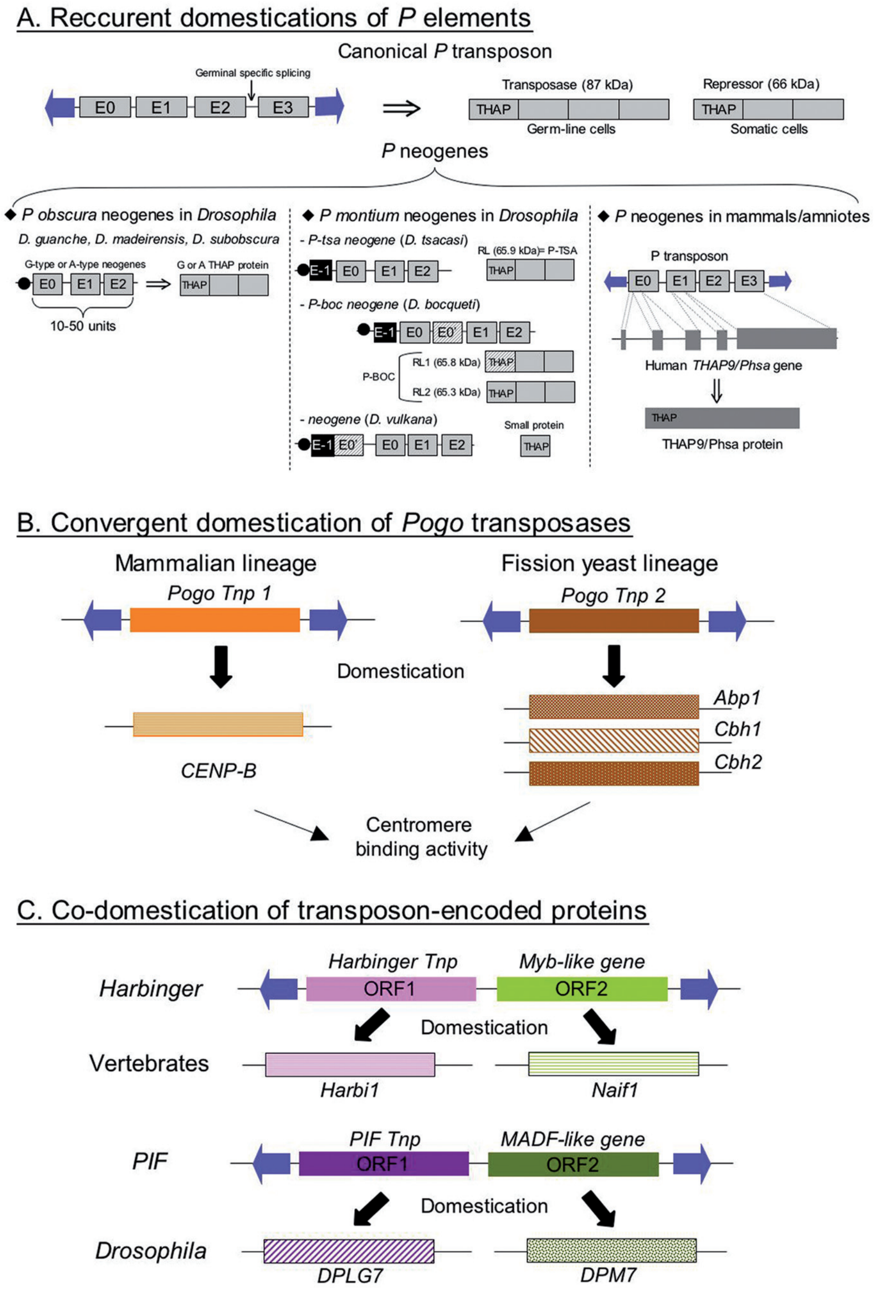

Figure 3. Evolutionary diversity in domestication events. (*A*) Recurrent domestication of *P* elements. The canonical *P* transposon produces a 87-kDa transposase in germline cells and a 66-kDa protein repressor in somatic cells in *Drosophila* through tissue-specific alternative splicing of intron 3. *P* elements underwent multiple domestication events in separate lineages of *Drosophila.* The promoter is represented by a black sphere. *P obscura* neogenes produce G or A THAP proteins. *P montium* neogenes comprise the *P-tsa* neogenes that produce a 65.9-kDa repressor-like protein (RL) and *P-boc* neogenes that arose from the acquisition of an untranslated exon (E-1) and an additional exon (E0). The alternative splicing of *P-boc* neogenes gives rise to two proteins, RL1 and RL2. The *D. vulkana* genome contains a third form of *P montium* neogenes, in which an additional exon (E0') is located upstream of exon E0. A domestication event of *P* elements had also occurred before the separation of birds and mammals that has led to a widespread occurrence of *P* neogenes in mammalian species, including humans. The coding region of a *P* transposase is compared with that of the human *THAP9/Phsa* gene. The THAP domains are a highly conserved feature of *P* neogenes in *Drosophila* as well as in mammals. (*B*) Convergent domestication of *Pogo* transposases. Distinct *Pogo* transposase sources independently gave rise to the CENP-B proteins in the mammalian lineage and to Abp1, Cbh1 and Cbh2, present in fission yeast. All of these proteins play roles associated with centromere-binding activity [126]. (*C*) Co-domestication of transposon-encoded proteins. *PIF/Harbinger* transposons encode two proteins: a transposase (Tnp) and a DNA-binding protein (referred to as Myb-like protein); both are required for transposition. The domestication process is associated with the immobilization of the two genes (by loss of the terminal inverted repeats) encoded by an ancestral active transposon. Two instances have been described in vertebrates and in *Drosophila.* HARBI1 and NAIF1 have emerged from a transposase and a Myb-like protein, respectively, encoded by an ancestral *Harbinger* transposon. DPLG7 and DPM7 originated from a transposase and a Myb-like gene encoded by an ancestral *PIF* transposon.

the primary transcript [105]. In the germline, the four exons (exons 0–3) give rise to an 87-kDa protein competent for genomic mobility (transposase). In somatic cells, the third intron is not spliced and a 66-kDa protein is produced, acting as a repressor of transposition [106]. The second class includes the different stationary forms of *P* element, including *obscura P* and *montium P* species subgroups of neogenes that became immobile through loss of the TIRs and the last exon required for transposase specificity (reviewed in [62]). The *obscura P* and *montium P* neogenes represent two distinct immobilization events of *P* transposons from the same ancestral *P* family [107, 108]. These neogenes have retained their coding capacity, and produce a protein similar to the 66-kDa *P* element repressor of *Drosophila melanogaster* (protein RL for repressor-like). The *obscura P* neogenes were originally found as repetitive units located at a single genomic site in the *Drosophila obscura* species subgroup that give rise to two proteins, G and A THAP proteins [26, 109]. The *montium P* neogenes are single-copy genes that occur in the *Drosophila montium* species subgroup and contain an untranslated new exon (Exon-1) [110, 111]. Different forms of *montium P* neogenes, the *P-tsa* and *P-boc* neogenes, evolved in several *Drosophila* species by capture of an additional exon 0 (referred to as exon 0') downstream of exon 0 [107]. In *Drosophila vulkana*, a similar exon-shuffling process gave rise to a neogene that contains an exon 0' located upstream of exon 0 [107, 111]. Products of the *P-tsa* and *P-boc* neogenes have been shown to bind chromatin *in vivo,* and do not repress transposition or transcription of canonical *P* element transposons [108]. These proteins are speculated to be involved in the regulation of the expression of many different euchromatic regions and/or in the modification of chromatin structure.

In addition to the immobilization of *P*-homologous sequences in *Diptera*, it has recently been found that a distinct domestication event of *P* elements had occurred before the separation of mammals and birds [112]. This molecular domestication event has led to a widespread occurrence of *P* neogenes in the vertebrate lineage including *Phsa* in human (Fig. 3A), *Pgga* in chicken and *Pdre* in zebrafish located at orthologous positions within their respective host genomes [65, 112, 113]. These genes as well as the *Diptera P* neogenes contain a THAP DBD demonstrating that the THAP domain is a recurrent theme in domestication of *P* elements [61, 112].

Convergent domestication

Many studies have pointed out the possible evolutionary relationship between CENP-B proteins and several *Pogo* transposases, including the *pogo* transposase from *D. melanogaster,* and the human *Tigger1* and *Tigger2* transposases [27, 114, 115]. Indeed, CENP-B and the *Pogo* transposase share striking sequence similarities through their DBDs as well as the [D, D, E/D] catalytic core [27]. Furthermore, the relationship is reinforced by the fact that the CENP-B box (the binding site of CENP-B) resembles the TIRs of *Tigger2* [115]. The CENP-B protein is highly conserved in mammalian species and has a central function in the assembly of centromere structure (reviewed in [116]). This protein binds specifically to the 17-bp CENP-B box located within highly repetitive alpha-satellite sequences positioned at the centromere of autosomes and the X chromosome [117, 118]. The N-terminal region of CENP-B proteins contains a DBD that forms one of the three groups distinguished within the Pipsqueak family proteins, whereas the C-terminal region mediates homodimerization [85, 117–119]. Three CENP-B homologs and their respective target sequences have been identified in the fission yeast *Schizosaccharomyces pombe:* ARS-binding protein (Abp1), CENP-B homolog 1 (cbh1) and CENP-B homolog 2 (Cbh2). These are involved in centromeric heterochromatin assembly, chromosome segregation and likely DNA replication initiation [120–124]. These proteins also play roles in retrotransposon silencing in yeast [125] (see 'Diversity in functional roles' below). Recently, Casola et al. [126] have proposed an evolutionary scenario of convergent domestication by which two distinct sources of *Pogo*-like transposase gave rise independently to mammalian CENP-B and the fission yeast proteins Abp1, Cbh1 and Cbh2 with centromere binding activity (Fig. 3B). Although the role of CENP-B in mammals is not clearly established, this protein is required for *de novo* centromere assembly on DNA lacking a functional centromere, and prevents the formation of excess centromeres on chromosomes [127, 128].

Co-domestication

PIF/Harbinger and *CACTA* form particular superfamilies of class II transposons in a sense that these elements encode two proteins that are necessary to mediate transposition [7–9]. Autonomous *PIF/Harbinger* transposons are characterized by two ORFs encoding a transposase and a Myb-like protein [99–101]. It was recently found that the Myb-like protein encoded by a resurrected zebrafish *Harbinger3_DR* transposon is required for transposition in at least two distinct functions: it promotes the nuclear import of the transposase, and recruits the transposase to the transposon ends [9].

Two examples of molecular domestication of *PIF/Harbinger* transposons in vertebrates and *Drosophila* species have been reported [9, 100, 101]. Casola et al. [101] have identified seven distinct transposase-derived genes called DPLG1–7 (*Drosophila PIF*-like genes 1–7) that probably arose from at least three independent domestication events. Furthermore, the authors have identified a domesticated Myb-like protein called DPMG7 (*Drosophila PIF* MADF-like protein-encoding gene 7) in a region close to the DPLG7A ortholog in three *Drosophila* species. These findings strongly support an evolutionary scenario of co-domestication by which the DPMG7 and DPLG7 genes have emerged from the same, formerly active *PIF/Harbinger* transposon (Fig. 3C).

We have recently described a similar co-domestication event involving two proteins conserved in bony vertebrates: HARBI1 (Harbinger derived-protein 1) evolved from a *Harbinger* transposase and NAIF1 that was identified as a protein containing a trihelix motif similar to that found in the myb-like protein [9, 100] (Fig. 3C). We have found that these two proteins have emerged from a common ancestor of jawed vertebrates after its separation from jawless vertebrates some 500 million years ago. The preliminary functional characterization of these two proteins have highlighted functional homologies with the transposase and the Myb-like protein of *Harbinger3_DR*, further supporting co-domestication. Indeed, similar to the interactions between the transposase and the Myb-like protein, NAIF1 interacts with HARBI1, promotes nuclear import of HARBI1 and acts as a DNA-binding protein [9]. Although NAIF1 and HARBI1 are speculated to be involved in the same molecular pathway, they are not yet functionally characterized.

Diversity in functional roles

Recruitment for a gene-regulatory function

Transposases carry two essential domains: a specific DBD and a catalytic domain responsible for DNA cleavage and joining reactions. Although a large number of domesticated proteins evolved from the entire transposase gene, and consequently contain both functional domains, DBDs appear to have preferentially been co-opted by the host. Thus, it is not surprising that the majority of domesticated proteins function as transcriptional regulators (activators or repressors) (Table 2). Recently, the human THAP7 protein containing a THAP DBD has been reported to display transcriptional regulatory properties *via* modification of chromatin structure [66]. THAP7 preferentially binds to hypoacetylated histone H4 tails, recruits the corepressors histone deacetylase (HDAC) 3 and the nuclear hormone receptor corepressor (NcoR) to promoters, and promotes histone H3 hypoacetylation [66]. Transcriptional regulation is not restricted to the THAP domain, but is also associated with the BED, the pipsqueak and the WRKY/GCM1 families of DBD. In human, the ZBED1 protein has been shown to act as a transcriptional activator of cell proliferation and ribosomal genes [80]. In *Saccharomyces cerevisiae*, the aft1 protein that contains a WRKY-type DBD is a transcription factor involved in ion utilization and homeostasis [129]. Psq has been shown to be essential for sequence-specific targeting of a Polycomb group complex that contains HDAC activity [130]. Psq binds specifically to the GAGA sequence that is present in many Hox genes and in hundreds of other chromosomal sites [130]. Furthermore, the recruitment of a transcriptional regulator may affect a variety of important biological processes, including DNA replication, morphogenesis, cell proliferation, growth and differentiation. However, the majority of domesticated proteins are incompletely characterized, and only hypothesized to function as transcription factors. An example for another gene regulatory mechanism is the Jerky protein that has emerged from a *Pogo* transposase, and is conserved in mammals. Jerky binds a large set of mRNAs and may regulate the availability of mRNAs to the translational machinery in neurons [93]. Moreover, Jerky-deficient mice develop epileptic seizures showing that this protein plays an important cellular role [92].

Chromatin-associated factors

Another prominent role played by domesticated proteins is the regulation of chromatin structure. The best-characterized proteins are the CENP-B proteins, BEAF-32 (described in ’Evolutionary diversity in domestication events’ above) and HIM-17. In *C. elegans*, HIM-17 is required for initiation of meiotic recombination, chromosome segregation and chiasma formation [131].

Apoptosis-related functions

Three domesticated proteins with apoptosis-related functions have been characterized. THAP0 (DAP4/p52rIPK) is involved in interferon γ-induced apoptosis in HeLa cells, and was identified as an activator of the interferon-induced protein kinase PKR, an important mediator of stress-induced apoptosis [132, 133]. The THAP1 protein is a nuclear proapoptotic factor that potentiates tumor necrosis factor α-induced apoptosis [134]. This protein is hypothesized to recruit the Par-4 protein (prostate-apoptosis-response-4) to specific promoters to stimulate or inhibit transcriptional activation of genes involved in apoptosis. Finally, the E93 protein is known to be a regulator of steroid-triggered programmed cell death during *Drosophila* development by playing an important role in activation of autophagic cell death [88, 135].

Cell-cycle control

Members of the THAP family of DBD have been shown to be involved in cell-cycle regulation. In *C. elegans*, the LIN-36 and LIN-15B proteins have been found to act as inhibitors of the G1/S transition [136]. The THAP-E2F6 fusion gene in fish species functions as a repressor of E2F-dependent transition during S phase that is critical for distinguishing G1/S and G2/M transcription during the cell cycle [137]. In human, the proapoptotic protein THAP1 was recently shown to be a regulator of endothelial cell proliferation and G1/S cell-cycle progression, which modulates expression of pRb (retinoblastoma)/E2F-dependent target genes including *RRM1* that is essential for S-phase DNA synthesis [138].

Capacity to mediate transposition

Domesticated genes evolved from components of active, mobile molecular parasites. In light of this consideration, it appears important to investigate whether transposase-related proteins could mobilize transposons in *trans* or act as transpositional regulators, which could affect host genome integrity. The evolutionary process of domestication is often accompanied by the modification of the [D, D, D/E] catalytic triad that is important for transposase activity. This change may not necessarily lead to the loss of transposase function, but suggests a modification of transposase activity that better suits the gain of a new host function. Some proteins have maintained a perfect [D, D, D/E] catalytic triad of amino acid residues such as *Buster* that was derived from a *hAT* transposon or HARBI1 [2, 100]. Nevertheless, the majority of those transposon-derived proteins that have been tested in transposition reaction were found defective. For example, the primate-specific SETMAR has preserved a specific DNA-binding ability of its ancestral transposase, but is defective in the DNA cleavage reaction that generates the 3'-hydroxyl group at the end of the transposon [53, 54]. However, the protein is fully active when supplied with precleaved transposon ends *in vitro*, suggesting that this protein could potentially mobilize *Hsmar1* transposons in the human genome [53]. Similar, the zebrafish ortholog of HARBI1 that emerged from a *Harbinger* transposase is deficient both in mediating transposition of *Harbinger3_DR* transposons (whose transposase is phylogenetically the closest to HARBI1) and in regulating transposition by the cognate transposase [9]. Kobuta, a domesticated protein derived from a *piggyBac* transposase in the *Xenopus* genome, was found inactive in *Uribo*-type *piggyBac* transposition [139]. Both the domesticated Kobuta protein and *Uribo* transposons coexist in the same genome.

The only transposon-derived elements shown to have retained the capacity to achieve transposition are the RAG1 protein together with its *cis*-regulatory RSS sequences, both probably evolved from a *Transib* transposon [35–38]. However, even though RAG-mediated transposition can be observed in cell-free reactions [32–34], transposition is an extremely rare event *in vivo* [36–38].

Roles to protect against transposon invasion

Given the potential of TEs to invade the host genome and cause detrimental effects, it is likely that host species have evolved different mechanisms to suppress or attenuate their activity. DNA transposons can become silenced *via* RNA interference (RNAi), a gene-silencing mechanism in which dsRNA triggers sequence-specific RNA degradation. This mechanism takes place in *C. elegans* to silence Tc1 transposition in the germline [140]. The authors have also shown that this silencing machinery can also suppress transposition of various unrelated transposons such as Tc3, Tc5, and Tc7. Silencing of TEs can also involve epigenetic modifications through post-translational modifications of histone tails and chromatin remodeling (reviewed in [141]). Transpositional activity can also be limited or repressed by the transposon itself *via* the production of a transpositional repressor; a good example for this is the *P* element repressor that is expressed in somatic tissues of *Drosophila* [106]. Second, overexpression of the transposase can reduce transposition activity *via* an overproduction-inhibition mechanism [142]. Finally, the production of a mutated transposase can antagonize the activity of the wild-type transposase through heterodimerization and dominant negative complementation (DNC) [142].

Even though transposon-derived proteins likely perform cellular functions that are not related to transposon regulation, several studies raise the question whether domesticated proteins could originally have been recruited as regulators or repressors of transposition by different processes including RNAi, epigenetic modifications or DNC. For example, it has been proposed that SETMAR could regulate *Hsmar1* transposase expression in human cells [54]. Similar, it has been proposed that the PGBD3 transposase was originally domesticated to repress the transposition of *piggyBac* and the associated non-autonomous element *MER85* [59]. Recently, DNA transposon-derived proteins have been shown to silence another class of TEs in *Schizosaccharomyces* *pombe* [125]. The Abp1 and Cbh1 proteins, which originated from a *Pogo* transposase, were initially known to act in centromeric heterochromatin formation and chromosome segregation in yeast [122, 123]. Cam et al. [125] have proposed that these two proteins have been co-opted by the host to control retroelement mobility. Abp1 may recruit Cbh2 and Cbh1 to *Tf2* retrotransposon long terminal repeats (LTRs) as well as LTR-associated genes. Abp1 negatively regulates *Tf2* expression by directly recruiting HDACs to *Tf2* and represses several genes through nearby LTRs (reviewed in [143]).

Concluding remarks and future directions

The numerous examples of transposon-derived genes detailed above provide evidence that TEs have the capacity to profoundly influence genome function. Molecular domestication has led to the emergence of new host genes that display important cellular functions including transcriptional regulation, chromatin-based control of the cell cycle, cell proliferation, apoptosis and chromatin structure. Despite the growing list of these genes, only few have been functionally characterized. Future investigations into the mechanisms and evolution of TEs will undoubtedly facilitate the discovery of new domesticated genes and their functional characterization. Due to the conservation of functional domain(s), some domesticated genes may have preserved some specific activities of the ancestral transposase. Thus, their biological roles can potentially be elucidated based on mechanistic similarities to *bona fide* transposition reactions. Moreover, in one recent report, three domesticated genes that exert crucial biological roles *in vivo* were also found to be involved in cellular mechanisms for silencing transposon activity [125]. Regulation of TE activity remains one of the most interesting aspects in current TE research, and we predict that other examples of domestication of transposon-encoded proteins for transposition control will be uncovered.

Acknowledgements. We thank V. Kapitonov and C. Feschotte for sharing with us their thoughts on transposon evolution and domestication.

1 Finnegan, D. J. (1989) Eukaryotic transposable elements and genome evolution. Trends Genet. 5, 103–107.

2 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K. et al. (2001) Initial sequencing and analysis of the human genome. Nature 409, 860–921.

3 Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P. and Schulman, A. H. (2007) A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8, 973–982.

4 Kapitonov, V. V. and Jurka, J. (2001) Rolling-circle transposons in eukaryotes. Proc. Natl. Acad. Sci. USA 98, 8714–8719.

5 Pritham, E. J., Putliwala, T. and Feschotte, C. (2007) Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene 390, 3–17.

6 Craig, N. L., Craigie, R., Gellert, M. and Lambowitz, A. M. (2002) Mobile DNA II. ASM Press, Washington, DC.

7 Frey, M., Reinecke, J., Grant, S., Saedler, H. and Gierl, A. (1990) Excision of the En/Spm transposable element of *Zea mays* requires two element-encoded proteins. EMBO J. 9, 4037–4044.

8 Yang, G., Zhang, F., Hancock, C. N. and Wessler, S. R. (2007) Transposition of the rice miniature inverted repeat transposable element mPing in *Arabidopsis thaliana.* Proc. Natl. Acad. Sci. USA 104, 10962–10967.

9 Sinzelle, L., Kapitonov, V. V., Grzela, D. P., Jursch, T., Jurka, J., Izsvák, Z. and Ivics, Z. (2008) Transposition of a reconstructed Harbinger element in human cells and functional homology with two transposon-derived cellular genes. Proc. Natl. Acad. Sci. USA 105, 4715–4720.

10 Lisch, D., Chomet, P. and Freeling, M. (1995) Genetic characterization of the Mutator system in maize: Behavior and regulation of Mu transposons in a minimal line. Genetics 139, 1777–1779.

11 Raizada, M. N. and Walbot, V. (2000) The late developmental pattern of Mu transposon excision is conferred by a cauliflower mosaic virus 35S-driven MURA cDNA in transgenic maize. Plant Cell 12, 5–21.

12 Lisch, D. (2002) Mutator transposons. Trends Plant Sci. 7, 498–504.

13 Doak, T. G., Doerder, F. P., Jahn, C. L. and Herrick, G. (1994) A proposed superfamily of transposase genes: Transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc. Natl. Acad. Sci. USA 91, 942–946.

14 Haren, L., Ton-Hoang, B. and Chandler, M. (1999) Integrating DNA: Transposases and retroviral integrases. Annu. Rev. Microbiol. 53, 245–281.

15 Kidwell, M. G. and Lisch, D. R. (2001) Perspective: Transposable elements, parasitic DNA, and genome evolution. Evolution 55, 1–24.

16 Bowen, N. J. and Jordan, I. K. (2002) Transposable elements and the evolution of eukaryotic complexity. Curr. Issues Mol. Biol. 4, 65–76.

17 Mariño-Ramírez, L., Lewis, K. C., Landsman, D. and Jordan, I. K. (2005) Transposable elements donate lineage-specific regulatory sequences to host genomes. Cytogenet. Genome Res. 110, 333–341.

18 Belenkaya, T., Soldatov, A., Nabirochkina, E., Birjukova, I., Georgieva, S. and Georgiev, P. (1998) P-Element insertion at the polyhomeotic gene leads to formation of a novel chimeric protein that negatively regulates yellow gene expression in P-element-induced alleles of *Drosophila melanogaster.* Genetics 150, 687–697.

19 Zhang, J., Zhang, F. and Peterson, T. (2006) Transposition of reversed Ac element ends generates novel chimeric genes in maize. PLoS Genet. 2, e164.

20 Jiang, N., Bao, Z., Zhang, X., Eddy, S. R. and Wessler, S. R. (2004) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431, 569–573.

21 Dooner, H. K. and Weil, C. F. (2007) Give-and-take: Interactions between DNA transposons and their host plant genomes. Curr. Opin. Genet. Dev. 17, 486–492.

22 Britten, R. (2006) Transposable elements have contributed to thousands of human proteins. Proc. Natl. Acad. Sci. USA 103, 1798–1803.

23 Wu, M., Li, L. and Sun, Z. (2006) Transposable element fragments in protein-coding regions and their contributions to human functional proteins. Gene 401, 165–171.

24 Nekrutenko, A. and Li, W. H. (2001) Transposable elements are found in a large number of human protein-coding genes. Trends Genet. 17, 619–621.

25 Li, W. H., Gu, Z., Wang, H. and Nekrutenko, A. (2001) Evolutionary analyses of the human genome. Nature 409, 847–849.

26 Miller, W. J., Hagemann, S., Reiter, E. and Pinsker, W. (1992) P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*. Proc. Natl. Acad. Sci. USA 89, 4018–4022.

27 Smit, A. F. and Riggs, A. D. (1996) Tiggers and DNA transposon fossils in the human genome. Proc. Natl. Acad. Sci. USA 93, 1443–1448.

28 Smit, A. F. (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. 9, 657–663.

29 Volff, J. N. (2006) Turning junk into gold: Domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28, 913–922.

30 Feschotte, C. and Pritham, E. J. (2007) DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41, 331–368.

31 Jones, J. M. and Gellert, M. (2004) The taming of a transposon: V(D)J recombination and the immune system. Immunol. Rev. 200, 233–248.

32 Agrawal, A., Eastman, Q. M. and Schatz, D. G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 394, 744–751.

33 Zhou, L., Mitra, R., Atkinson, P. W., Hickman, A. B., Dyda, F. and Craig, N. L. (2004) Transposition of hAT elements links transposable elements and V(D)J recombination. Nature 432, 995–1001.

34 Hiom, K., Melek, M. and Gellert, M. (1998) DNA transposition by the RAG1 and RAG2 proteins: A possible source of oncogenic translocations. Cell 94, 463–470.

35 Clatworthy, A. E., Valencia, M. A., Haber, J. E. and Oettinger, M. A. (2003) V(D)J recombination and RAG-mediated transposition in yeast. Mol. Cell 12, 489–499.

36 Chatterji, M., Tsai, C. L. and Schatz, D. G. (2006) Mobilization of RAG-generated signal ends by transposition and insertion *in vivo*. Mol. Cell. Biol. 26, 1558–1568.

37 Reddy, Y. V., Perkins, E. J. and Ramsden, D. A. (2006) Genomic instability due to V(D)J recombination-associated transposition. Genes Dev. 20, 1575–1582.

38 Kapitonov, V. V. and Jurka, J. (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. 3, e181.

39 Panchin, Y. and Moroz, L. L. (2008) Molluscan mobile elements similar to the vertebrate recombination-activating genes. Biochem. Biophys. Res. Commun. 369, 818–823.

40 Hudson, M. E., Lisch, D. R. and Quail, P. H. (2003) The FHY3 and FAR1 genes encode transposase-related proteins involved in regulation of gene expression by the phytochrome A-signaling pathway. Plant J. 34, 453–471.

41 Lin, R., Ding, L., Casola, C., Ripoll, D. R., Feschotte, C. and Wang, H. (2007) Transposase-derived transcription factors regulate light signaling in *Arabidopsis*. Science 318, 1302–1305.

42 Babu, M. M., Iyer, L. M., Balaji, S. and Aravind, L. (2006) The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. Nucleic Acids Res. 34, 6505–6520.

43 Barkan, A. and Martienssen, R. A. (1991) Inactivation of maize transposon Mu suppresses a mutant phenotype by activating an outward-reading promoter near the end of Mu1. Proc. Natl. Acad. Sci. USA 88, 3502–3506.

44 Benito, M. I. and Walbot, V. (1997) Characterization of the maize mutator transposable element MURA transposase as a DNA-binding protein. Mol. Cell. Biol. 17, 5165–5175.

45 Raizada, M. N., Benito, M. I. and Walbot, V. (2001) The MuDR transposon terminal inverted repeat contains a complex plant promoter directing distinct somatic and germinal programs. Plant J. 25, 79–91.

46 Xu, Z., Yan, X., Maurais, S., Fu, H., O'Brien, D. G., Mottinger, J. and Dooner, H. K. (2004) Jittery, a mutator distant relative with a paradoxical mobile behavior: Excision without reinsertion. Plant Cell 16, 1105–1114.

47 Cowan, R. K., Hoen, D. R., Schoen, D. J. and Bureau, T. E. (2005) MUSTANG is a novel family of domesticated transposase genes found in diverse angiosperms. Mol. Biol. Evol. 22, 2084–2089.

48 Bundock, P. and Hooykaas, P. (2005) An *Arabidopsis* hAT-like transposase is essential for plant development. Nature 436, 282–284.

49 Muehlbauer, G. J., Bhau, B. S., Syed, N. H., Heinen, S., Cho, S., Marshall, D., Pateyron, S., Buisine, N., Chalhoub, B. and Flavell, A. J. (2006) A hAT superfamily transposase recruited by the cereal grass genome. Mol. Genet. Genomics 275, 553–563.

50 Robertson, H. M. and Zumpano, K. L. (1997) Molecular evolution of an ancient mariner transposon, Hsmar1, in the human genome. Gene 205, 203–217.

51 Cordaux, R., Udit, S., Batzer, M. A. and Feschotte, C. (2006) Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc. Natl. Acad. Sci. USA 103, 8101–8106.

52 Lee, S. H., Oshige, M., Durant, S. T., Rasila, K. K., Williamson, E. A., Ramsey, H., Kwan, L., Nickoloff, J. A., Hromas, R. (2005) The SET domain protein Metnase mediates foreign DNA integration and links integration to nonhomologous end-joining repair. Proc. Natl. Acad. Sci. USA 102, 18075–18080.

53 Liu, D., Bischerour, J., Siddique, A., Buisine, N., Bigot, Y. and Chalmers, R. (2007) The human SETMAR protein preserves most of the activities of the ancestral Hsmar1 transposase. Mol. Cell. Biol. 27, 1125–1132.

54 Miskey, C., Papp, B., Mátés, L., Sinzelle, L., Keller, H., Izsvák, Z. and Ivics, Z. (2007) The ancient mariner sails again: Transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. Mol. Cell. Biol. 27, 4589–4600.

55 Roman, Y., Oshige, M., Lee, Y. J., Goodwin, K., Georgiadis, M. M., Hromas, R. A. and Lee, S. H. (2007) Biochemical characterization of a SET and transposase fusion protein, Metnase: Its DNA binding and DNA cleavage activity. Biochemistry 46, 11369–11376.

56 Beck, B. D., Park, S. J., Lee, Y. J., Roman, Y., Hromas, R. A. and Lee, S. H. (2008) Human PSO4 is a Metnase (SETMAR) binding partner that regulates Metnase' function in DNA repair. J. Biol. Chem. 283, 9023–9030.

57 Sarkar, A., Sim, C., Hong, Y. S., Hogan, J. R., Fraser, M. J., Robertson, H. M. and Collins, F. H. (2003) Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. Mol. Genet. Genomics 270, 173–180.

58 Edelstein, L. C. and Collins, T. (2005) The SCAN domain family of zinc finger transcription factors. Gene 359, 1–17.

59 Newman, J. C., Bailey, A. D., Fan, H. Y., Pavelitz, T. and Weiner, A. M. (2008) An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome. PLoS Genet. 4, e1000031.

60 Tipney, H. J., Hinsley, T. A., Brass, A., Metcalfe, K., Donnai, D. and Tassabehji, M. (2004) Isolation and characterisation of GTF2IRD2, a novel fusion gene and member of the TFII-I family of transcription factors, deleted in Williams-Beuren syndrome. Eur. J. Hum. Genet. 12, 551–560.

61 Roussigne, M., Kossida, S., Lavigne, A. C., Clouaire, T., Ecochard, V., Glories, A., Amalric, F. and Girard, J. P. (2003) The THAP domain: A novel protein motif with similarity to the DNA-binding domain of P element transposase. Trends Biochem. Sci. 28, 66–69.

62 Quesneville, H., Nouaud, D. and Anxolabehere, D. (2005) Recurrent recruitment of the THAP DNA-binding domain

and molecular domestication of the P-transposable element. Mol. Biol. Evol. 22, 741–746.

63 Clouaire, T., Roussigne, M., Ecochard, V., Mathe, C., Amalric, F. and Girard, J. P. (2005) The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. Proc. Natl. Acad. Sci. USA 102, 6907–6912.

64 Lee, C. C., Beall, E. L. and Rio, D. C. (1998) DNA binding by the KP repressor protein inhibits P-element transposase activity *in vitro.* EMBO J. 17, 4166–4174.

65 Hagemann, S. and Pinsker, W. (2001) *Drosophila* P transposons in the human genome? Mol. Biol. Evol. 18, 1979–1982.

66 Macfarlan, T., Kutney, S., Altman, B., Montross, R., Yu, J. and Chakravarti, D. (2005) Human THAP7 is a chromatin-associated, histone tail-binding protein that represses transcription *via* recruitment of HDAC3 and nuclear hormone receptor corepressor. J. Biol. Chem. 280, 7346–7358.

67 Chesney, M. A., Kidd, A. R. 3rd and Kimble, J. (2006) gon-14 functions with class B and class C synthetic multivulva genes to control larval growth in *Caenorhabditis elegans.* Genetics 172, 915–928.

68 Bessière, D., Lacroix, C., Campagne, S., Ecochard, V., Guillet, V., Mourey, L., Lopez, F., Czaplicki, J., Demange, P., Milon, A., Girard, J. P. and Gervais V. (2008) Structure-function analysis of the THAP zinc finger of THAP1, a large C2CH DNA-binding module linked to Rb/E2F pathways. J. Biol. Chem. 283, 4352–4363.

69 Aravind, L. (2000) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem. Sci. 25, 421–423.

70 Zhao, K., Hart, C. M. and Laemmli, U. K. (1995) Visualization of chromosomal domains with boundary element-associated factor BEAF-32. Cell 81, 879–889.

71 Hirose, F., Yamaguchi, M., Kuroda, K., Omori, A., Hachiya, T., Ikeda, M., Nishimoto, Y. and Matsukage, A. (1996) Isolation and characterization of cDNA for DREF, a promoter-activating factor for *Drosophila* DNA replication-related genes. J. Biol. Chem. 271, 3930–3937.

72 Hart, C. M., Zhao, K. and Laemmli, U. K. (1997) The scs' boundary element: Characterization of boundary element-associated factors. Mol. Cell. Biol. 17, 999–1009.

73 Hart, C. M., Cuvier, O. and Laemmli, U. K. (1999) Evidence for an antagonistic relationship between the boundary element-associated factor BEAF and the transcription factor DREF. Chromosoma 108, 375–383.

74 Yoshida, H., Inoue, Y. H., Hirose, F., Sakaguchi, K., Matsukage, A. and Yamaguchi, M. (2001) Over-expression of DREF in the *Drosophila* wing imaginal disc induces apoptosis and a notching wing phenotype. Genes Cells 6, 877–886.

75 Cuvier, O., Hart, C. M. and Laemmli, U. K. (1998) Identification of a class of chromatin boundary elements. Mol. Cell. Biol. 18, 7478–7486.

76 Gilbert, M. K., Tan, Y. Y. and Hart, C. M. (2006) The *Drosophila* boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. Genetics 173, 1365–1375.

77 Hirose, F., Ohshima, N., Shiraki, M., Inoue, Y. H., Taguchi, O., Nishi, Y., Matsukage, A. and Yamaguchi, M. (2001) Ectopic expression of DREF induces DNA synthesis, apoptosis, and unusual morphogenesis in the Drosophila eye imaginal disc: Possible interaction with Polycomb and trithorax group proteins. Mol. Cell. Biol. 21, 7231–7242.

78 Ohshima, N., Takahashi, M. and Hirose, F. (2003) Identification of a human homologue of the DREF transcription factor with a potential role in regulation of the histone H1 gene. J. Biol. Chem. 278, 22928–22938.

79 Esposito, T., Gianfrancesco, F., Ciccodicola, A., Montanini, L., Mumm, S., D'Urso, M. and Forabosco, A. (1999) A novel pseudoautosomal human gene encodes a putative protein similar to Ac-like transposases. Hum. Mol. Genet. 8, 61–67.

80 Yamashita, D., Sano, Y., Adachi, Y., Okamoto, Y., Osada, H., Takahashi, T., Yamaguchi, T., Osumi, T. and Hirose, F. (2007) hDREF regulates cell proliferation and expression of ribosomal protein genes. Mol. Cell. Biol. 27, 2003–2013.

81 Essers, L., Adolphs, R. H. and Kunze, R. (2000) A highly conserved domain of the maize activator transposase is involved in dimerization. Plant Cell 12, 211–224.

82 Michel, K., O'Brochta, D. A. and Atkinson, P. W. (2003) The C-terminus of the Hermes transposase contains a protein multimerization domain. Insect Biochem. Mol. Biol. 33, 959–970.

83 Yamashita, D., Komori, H., Higuchi, Y., Yamaguchi, T., Osumi, T. and Hirose, F. (2007) Human DNA replication-related element binding factor (hDREF) self-association via hATC domain is necessary for its nuclear accumulation and DNA binding. J. Biol. Chem. 282, 7563–7575.

84 Breitling, R. and Gerber, J. K. (2000) Origin of the paired domain. Dev. Genes Evol. 210, 644–650.

85 Siegmund, T. and Lehmann, M. (2002) The *Drosophila* Pipsqueak protein defines a new family of helix-turn-helix DNA-binding proteins. Dev. Genes Evol. 212, 152–157.

86 Godt, D., Couderc, J. L., Cramton, S. E. and Laski, F. A. (1993) Pattern formation in the limbs of *Drosophila:* Bric à brac is expressed in both a gradient and a wave-like pattern and is required for specification and proper segmentation of the tarsus. Development 119, 799–812.

87 Lehmann M, Siegmund T, Lintermann KG, Korge G (1998) The pipsqueak protein of *Drosophila melanogaster* binds to GAGA sequences through a novel DNA-binding domain. J. Biol. Chem. 1998, 273, 28504–28509.

88 Lee, C. Y., Wendel, D. P., Reid, P., Lam, G., Thummel, C. S. and Baehrecke, E. H. (2000) E93 directs steroid-triggered programmed cell death in *Drosophila.* Mol. Cell 6, 433–443.

89 Wang, H., Hartswood, E. and Finnegan, D. J. (1999) Pogo transposase contains a putative helix-turn-helix DNA binding domain that recognises a 12 bp sequence within the terminal inverted repeats. Nucleic Acids Res. 27, 455–461.

90 Robertson, H. M. (2002) Evolution of DNA transposon in eukaryotes. In: Mobile DNA II, pp. 1093–1110, Craig, N., L., Craigie, R., Gellert, M. and Lambowitz, A. M. (eds.), ASM Press, Washington, DC.

91 Dou, T., Gu, S., Zhou, Z., Ji, C., Zeng, L., Ye, X., Xu, J., Ying, K., Xie, Y. and Mao, Y. (2004) Isolation and characterization of a Jerky and JRK/JH8 like gene, tigger transposable element derived 7, TIGD7. Biochem. Genet. 42, 279–285.

92 Toth, M., Grimsby, J., Buzsaki, G. and Donovan, G. P. (1995) Epileptic seizures caused by inactivation of a novel gene, jerky, related to centromere binding protein-B in transgenic mice. Nat. Genet. 11, 71–75.

93 Liu, W., Seto, J., Sibille, E. and Toth, M. (2003) The RNA binding domain of Jerky consists of tandemly arranged helix-turn-helix/homeodomain-like motifs and binds specific sets of mRNAs. Mol. Cell. Biol. 23, 4083–4093.

94 Zeng, Z., Kyaw, H., Gakenheimer,K. R., Augustus, M., Fan, P., Zhang, X., Su, K., Carter, K. and Li, Y. (1997) Cloning, mapping, and tissue distribution of a human homologue of the mouse jerky gene product. Biochem. Biophys. Res. Commun. 236, 389–395.

95 Gabrielsen, O. S., Sentenac, A. and Fromageot, P. (1991) Specific DNA binding by c-Myb: Evidence for a double helix-turn-helix-related motif. Science 253, 1140–1143.

96 Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S. and Nishimura, Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell 79, 639–648.

97 Aasland, R., Stewart, A. F. and Gibson, T. (1996) The SANT domain: A putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. Trends Biochem. Sci. 21, 87–88.

98 Boyer, L. A., Latek, R. R. and Peterson, C. L. (2004) The SANT domain: A unique histone-tail-binding module? Nat. Rev. Mol. Cell Biol. 5, 158–163.

99 Zhang, X., Jiang, N., Feschotte, C. and Wessler, S. R. (2004) PIF- and Pong-like transposable elements: Distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. Genetics 166, 971–986.

100 Kapitonov, V. V. and Jurka, J. (2004) Harbinger transposons and an ancient HARBI1 gene derived from a transposase. DNA Cell Biol. 23, 311–324.

101 Casola, C., Lawing, A. M., Betrán, E. and Feschotte, C. (2007) PIF-like transposons are common in drosophila and have been repeatedly domesticated to generate new host genes. Mol. Biol. Evol. 24, 1872–1888.

102 Lv, B., Shi, T., Wang, X., Song, Q., Zhang, Y., Shen, Y., Ma, D. and Lou, Y. (2006) Overexpression of the novel human gene, nuclear apoptosis-inducing factor 1, induces apoptosis. Int. J. Biochem. Cell Biol. 38, 671–683.

103 Miller, W. J. and Capy, P. (2006) Applying mobile genetic elements for genome analysis and evolution. Mol. Biotechnol. 33, 161–174.

104 Miller, W. J., McDonald, J. F., Nouaud, D. and Anxolabéhère, D. (1999) Molecular domestication – More than a sporadic episode in evolution. Genetica 107, 197–207.

105 Laski, F. A., Rio, D. C. and Rubin, G. M. (1986) Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. Cell 44, 7–19.

106 Rio, D. C., Laski, F. A. and Rubin, G. M. (1986) Identification and immunochemical analysis of biologically active *Drosophila* P element transposase. Cell 44, 21–32.

107 Nouaud, D., Quesneville, H., Anxolabéhère, D. (2003) Recurrent exon shuffling between distant P-element families. Mol. Biol. Evol. 20, 190–199.

108 Reiss, D., Nouaud, D., Ronsseray, S. and Anxolabéhère, D. (2005) Domesticated P elements in the *Drosophila montium* species subgroup have a new function related to a DNA binding property. J. Mol. Evol. 61, 470–480.

109 Miller W., Paricio, N., Hagemann, S., Martínez-Sebastián, M. J., Pinsker, W., de Frutos, R. (1995) Structure and expression of clustered P element homologues in *Drosophila subobscura* and *Drosophila guanche.* Gene 156, 167–174.

110 Nouaud, D. and Anxolabéhère, D. (1997) P element domestication: A stationary truncated P element may encode a 66-kDa repressor-like protein in the *Drosophila montium* species subgroup. Mol. Biol. Evol. 14, 1132–1144.

111 Nouaud, D., Boëda, B., Levy, L. and Anxolabéhère, D. (1999) A P element has induced intron formation in *Drosophila.* Mol. Biol. Evol. 16, 1503–1510.

112 Hammer, S. E., Strehl, S. and Hagemann, S. (2005) Homologs of *Drosophila* P transposons were mobile in zebrafish but have been domesticated in a common ancestor of chicken and human. Mol. Biol. Evol. 22, 833–844.

113 Hagemann, S. and Hammer, S. E. (2006) The implications of DNA transposons in the evolution of P elements in zebrafish (*Danio rerio*). Genomics. 88, 572–579.

114 Tudor, M., Lobocka, M., Goodell, M., Pettitt, J. and O'Hare, K. (1992) The pogo transposable element family of *Drosophila melanogaster.* Mol. Gen. Genet. 232, 126–134.

115 Kipling, D. and Warburton, P. E. (1997) Centromeres, CENP-B and Tigger too. Trends Genet. 13, 141–145.

116 Masumoto, H., Nakano, M. and Ohzeki, J. (2004) The role of CENP-B and alpha-satellite DNA: *De novo* assembly and epigenetic maintenance of human centromeres. Chromosome Res. 12, 543–556.

117 Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J. Cell Biol. 109, 1963–1973.

118 Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M. and Okazaki, T. (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. J. Cell Biol. 116, 585–596.

119 Yoda, K., Kitagawa, K., Masumoto, H., Muro, Y. and Okazaki, T. (1992) A human centromere protein, CENP-B, has a DNA binding domain containing four potential alpha helices at the NH_2 terminus, which is separable from dimerizing activity. J. Cell Biol. 119, 1413–1427.

120 Murakami, Y., Huberman, J. A. and Hurwitz, J. (1996) Identification, purification, and molecular cloning of autonomously replicating sequence-binding protein 1 from fission yeast *Schizosaccharomyces pombe.* Proc. Natl. Acad. Sci. USA 93, 502–507.

121 Lee, J. K., Huberman, J. A. and Hurwitz, J. (1997) Purification and characterization of a CENP-B homologue protein that binds to the centromeric K-type repeat DNA of *Schizosaccharomyces pombe.* Proc. Natl. Acad. Sci. USA 94, 8427–8432.

122 Irelan, J. T., Gutkin, G. I. and Clarke, L. (2001) Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. Genetics 157, 1191–1203.

123 Nakagawa, H., Lee, J. K., Hurwitz, J., Allshire, R. C., Nakayama, J., Grewal, S. I., Tanaka, K. and Murakami, Y. (2002) Fission yeast CENP-B homologs nucleate centromeric heterochromatin by promoting heterochromatin-specific histone tail modifications. Genes Dev. 16, 1766–1778.

124 Locovei, A. M., Spiga, M. G., Tanaka, K., Murakami, Y. and D'Urso, G. (2006) The CENP-B homolog, Abp1, interacts with the initiation protein Cdc23 (MCM10) and is required for efficient DNA replication in fission yeast. Cell Div. 1, 27.

125 Cam, H. P., Noma, K., Ebina, H., Levin, H. L. and Grewal, S. I. (2008) Host genome surveillance for retrotransposons by transposon-derived proteins. Nature 451, 431–436.

126 Casola, C., Hucks, D. and Feschotte, C. (2008) Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. Mol. Biol. Evol. 25, 29–41.

127 Ohzeki, J., Nakano, M., Okada, T. and Masumoto, H. (2002) CENP-B box is required for *de novo* centromere chromatin assembly on human alphoid DNA. J. Cell Biol. 159, 765–775.

128 Okada, T., Ohzeki, J., Nakano, M., Yoda, K., Brinkley, W. R., Larionov, V. and Masumoto, H. (2007) CENP-B controls centromere formation depending on the chromatin context. Cell 131, 1287–1300.

129 Yamaguchi-Iwai, Y., Dancis, A. and Klausner, R. D. (1995) AFT1: A mediator of iron regulated transcriptional control in *Saccharomyces cerevisiae.* EMBO J. 14, 1231–1239.

130 Ringrose, L., Ehret, H. and Paro, R. (2004) Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol. Cell. 16, 641–653.

131 Reddy, K. C. and Villeneuve, A. M. (2004) *C. elegans* HIM-17 links chromatin modification and competence for initiation of meiotic recombination. Cell 118, 439–452.

132 Deiss, L. P., Feinstein, E., Berissi, H., Cohen, O. and Kimchi, A. (1995) Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. Genes Dev. 9, 15–30.

133 Gale, M. Jr., Blakely, C. M., Hopkins, D. A., Melville, M. W., Wambach, M., Romano, P. R. and Katze, M. G. (1998) Regulation of interferon-induced protein kinase PKR: Modulation of P58IPK inhibitory function by a novel protein, P52rIPK. Mol. Cell. Biol. 18, 859–871.

134 Roussigne, M., Cayrol, C., Clouaire, T., Amalric, F. and Girard, J. P. (2003) THAP1 is a nuclear proapoptotic factor that links prostate-apoptosis-response-4 (Par-4) to PML nuclear bodies. Oncogene 22, 2432–2442.

135 Lee, C. Y. and Baehrecke,E. H. (2001) Steroid regulation of autophagic programmed cell death during development. Development 128, 1443–1455.

136 Boxem, M. and van den Heuvel, S. (2002) *C. elegans* class B synthetic multivulva genes act in G(1) regulation. Curr. Biol. 12, 906–911.

137 Giangrande, P. H., Zhu, W., Schlisio, S., Sun, X., Mori, S., Gaubatz, S. and Nevins, J. R. (2004) A role for E2F6 in distinguishing G1/S- and G2/M-specific transcription. Genes Dev. 18, 2941–2951.

138 Cayrol, C., Lacroix, C., Mathe, C., Ecochard, V., Ceribelli, M., Loreau, E., Lazar, V., Dessen, P., Mantovani, R., Aguilar, L. and Girard, J. P. (2007) The THAP-zinc finger protein THAP1 regulates endothelial cell proliferation through modulation of pRB/E2F cell-cycle target genes. Blood 109, 584–594.

139 Hikosaka, A., Kobayashi, T., Saito, Y. and Kawahara, A. (2007) Evolution of the *Xenopus* piggyBac transposon family TxpB: Domesticated and untamed strategies of transposon subfamilies. Mol. Biol. Evol. 24, 2648–2656.

140 Sijen, T. and Plasterk, R. H. (2003) Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. Nature 426, 310–314.

141 Slotkin, R. K. and Martienssen, R. (2007) Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. 8, 272–85.

142 Lohe, A. R. and Hartl, D. L. (1996) Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol. Biol. Evol. 13, 549–555.

143 O'Donnell, K. A. and Boeke, J. D. (2008) Domesticated DNA transposon proteins mediate retrotransposon control. Cell Res. 18, 331–333.

144 Lours, C., Bardot, O., Godt, D., Laski, F. A. and Couderc, J. L. (2003) The *Drosophila melanogaster* BTB proteins bric à brac bind DNA through a composite DNA binding domain containing a pipsqueak and an AT-Hook motif. Nucleic Acids Res. 31, 5389–5398.

145 Mojzita, D. and Hohrmann, S. (2006) Pdc2 coordinates expression of the THI regulon in the yeast *Saccharomyces cerevisiae.* Mol. Genet. Genomics 276, 147–161.

146 Gil, R., Zueco, J., Sentandreu, R. and Herrero, E. (1991) RCS1, a gene involved in controlling cell size in *Saccharomyces cerevisiae.* Yeast 7, 1–14.

147 Ishii, N., Yamamoto, M., Yoshihara, F., Arisawa, M. and Aoki, Y. (1997) Biochemical and genetic characterization of Rbf1p, a putative transcription factor of *Candida albicans.* Microbiology 143, 429–435.

148 Saito, R. M., Perreault, A., Peach, B., Satterlee, J. S. and van den Heuvel, S. (2004) The CDC-14 phosphatase controls developmental cell-cycle arrest in *C. elegans.* Nat. Cell Biol. 6, 777–783.

149 Chinnadurai, G. (2002) CtBP, an unconventional transcriptional corepressor in development and oncogenesis. Mol. Cell 9, 213–224.

150 Fay, D. S., Keenan, S. and Han, M. (2002) fzr-1 and lin-35/Rb function redundantly to control cell proliferation in *C. elegans* as revealed by a nonbiased synthetic screen. Genes Dev. 16, 503–517.

151 Roccaro, M., Li, Y., Sommer, H., Saedler, H. (2007) ROSINA (RSI) is part of a CACTA transposable element, TamRSI, and links flower development to transposon activity. Mol. Genet. Genomics 278, 243–254.

To access this journal online:
http://www.birkhauser.ch/CMLS