ARTICLE IN PRESS

Plasmid xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Plasmid

journal homepage: www.elsevier.com/locate/yplas

Review

Transposable elements domesticated and neofunctionalized by eukaryotic genomes

Ahmed M. Alzohairy [a], Gábor Gyulai [b], Robert K. Jansen [c,e], Ahmed Bahieldin [d,e,*]

[a] Genetics Department, Faculty of Agriculture, Zagazig University, Zagazig 44511, Egypt
[b] Institute of Genetics and Biotechnology, St. István University, Gödöllő H-2103, Hungary
[c] Section of Integrative Biology, University of Texas at Austin, Austin, TX 78712, USA
[d] Genetics Department, Faculty of Agriculture, Ain Shams University, Cairo, Egypt
[e] King Abdulaziz University, Faculty of Science, Department of Biological Sciences, Genomics and Biotechnology Section, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Article history:
Received 20 June 2012
Accepted 8 August 2012
Available online xxxx

Communicated by Prof. R. Chalmers

Keywords:
Transposable elements
Retrotransposons
DNA transposons
Gene domestication
Speciation

ABSTRACT

Whole genome sequencing has provided a massive amount of information about the origin, diversity and genomic impact of repetitive DNA sequences (repDNA). Among the many classes of repDNA, prokaryotic transposable elements (TEs) replicate, move, amplify and accumulate in invaded genomes and thus represent the major force in restructuring host genes and genomes during evolution. Similar to retroviruses, autonomous TEs became part of the host genomes, and after their molecular domestication, they became functional genes (genomic fossils) in eukaryotic genomes. In this review, examples of the domestication events are discussed, some of which are known to be induced by biotic and abiotic stressors.

Contents

* Corresponding author at: King Abdulaziz University, Faculty of Science, Department of Biological Sciences, Genomics and Biotechnology Section, Jeddah 21589, Saudi Arabia.

E-mail addresses: amansour@zu.edu.eg (A.M. Alzohairy), gyulai.gabor@mkk.szie.hu (G. Gyulai), jansen@austin.utexas.edu (R.K. Jansen), bahieldin55@gmail.com (A. Bahieldin).

0147-619X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.plasmid.2012.08.001

1. Introduction

Repetitive DNA is classified into high, middle and low copy repeats (Britten and Kohne, 1968), which include short to long repeats in genomes in a tandem, interspersed, or segmental organization (Sharp et al., 2006). Interspersed repeats of 20–30 kbp were found to be incomplete copies of transposable elements (TEs) of both DNA and retrotransposons (Debarry et al., 2006). All retrotransposons have two genes in common: *gag*, which encodes proteins (GAGs; *Group-specific AntiGen*) that form VLPs (*Virus-Like Particles*), and *pol* (*polymerase*), which encodes the three enzymes reverse transcriptase (RT), integrase (IN) and protease (PR) (Hou et al., 2010). The ability of TEs to change location within genomes results in new structural and functional activities (Wessler, 1996; Lisch and Bennetzen, 2011). These activities facilitate genetic changes that impact genome evolution and speciation (McClintock, 1984). The formation of new host genes through the recruitment of transposase functional domains supports the idea that transposable elements may flourish because they benefit their host (Britten, 2004; Volff, 2006; Casola et al., 2007; Feschotte, 2008).

Approximately 45% of the human genome is made up of TEs, many of which are inactive and ancient (Lander et al., 2001; Blumenstiel, 2011; Fig. 1). Plant genomes contain thousands of TEs that outnumber the genes and form the vast majority of the total DNA content (Feschotte et al., 2002; Schulman and Kalendar, 2005; Figs. 1 and 2). Similar to genomic fossils of viruses in eukaryotic genomes (Belyi et al., 2010), autonomous TEs have the ability to parasitize host genomes (Doolittle and Sapienza, 1980; Orgel and Crick, 1980; Fischer et al., 2000; Hughes and Coffin, 2001; Kazazian, 2004; Volff, 2006). When incorporated, TEs generate various types of genomic rearrangements through transpositions, which lead to gene insertions, deletions, duplications and inversions. When transposition occurs in coding and regulatory regions, it results in loss of genes or changes in the level of gene expression (Kashkush et al., 2003). Genes of TEs have been recruited (domesticated) by the host genome during evolution (Lyon, 2000; Nekrutenko and Li, 2001; Curcio and Derbyshire, 2003; Gotea and Makalowski, 2006). Such events are evident because there are a number of functional differences between autonomous ancestral TEs and domesticated TEs (Table 1). In plants, the genes *far*1, *fhy*3 and *frs* were domesticated from DNA transposons of Mutator-like elements (MULEs) (Cowan et al., 2005).

Approximately 50 to 100 protein-coding genes of mammalian and plant genomes evolved from coding sequences of TEs of both DNA transposons and retrotransposons (Lander et al., 2001; Kapitonov et al., 2004; Kapitonov and Jurka, 2004; Brandt et al., 2005; Kapitonov and Jurka, 2005; Campillos et al., 2006; Volff, 2006; Jurka et al.,

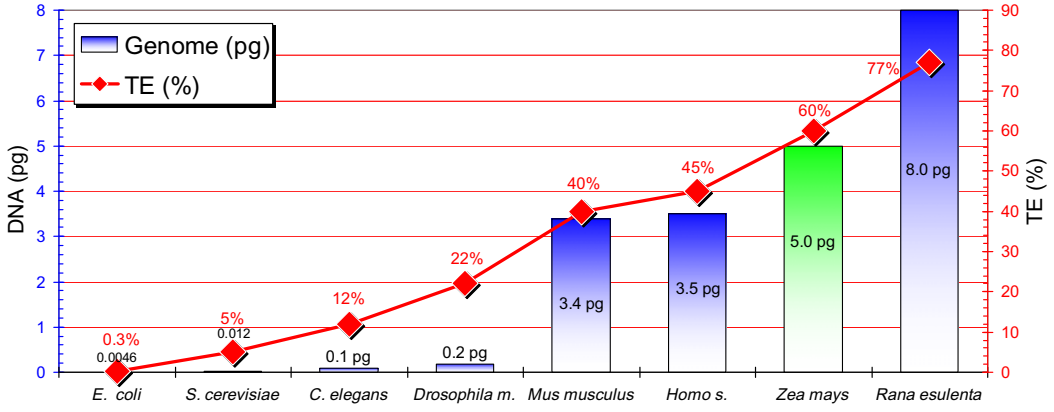

Fig. 1. Correlation between genome sizes DNA picograms (pg) and percentages of transposable elements (TE%) of species at different stages of the evolution. Data from Biémont and Vieira (2006) were edited and analyzed.

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001

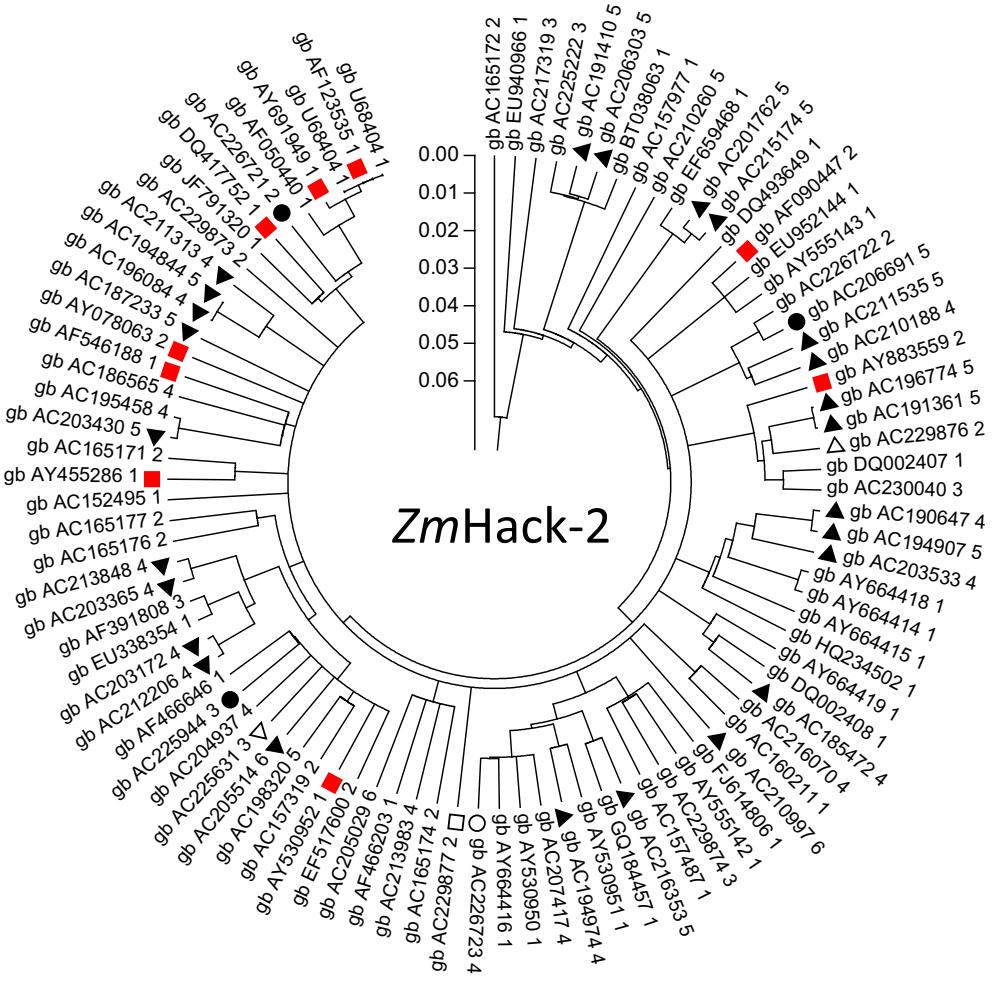

Fig. 2. Bootstrap consensus Neighbor Joining tree (NCBI Altschul et al., 1997) derived from the nucleotide sequences of the *ZmHack*-2 (SanMiguel et al., 1996), a *gypsy*-type plant LTR retrotransposon in *Zea mays* genome with extremely high copy number (up to 200,000) and sequence diversity (Feschotte et al., 2002). Sequence of the 5′LTR (1624 bp) of *Hack*-2 (U68404) was downloaded, blasted (NCBI; Altschul et al., 1997) and analyzed by Mega4 (Tamura et al., 2007). Nine loci, where *Hack*-2 jumped into a functional gene are labeled (■): genes of *alcohol dehydrogenase* (*adh*1, AF123535.1; and *adh*1A, AY691949.1); putative *heme oxygenase* (AY530952.1); *glume architecture*-1 gene (*tga*1, AY883559.2); the 19-kDa *zein* gene (AF546188.1); the nuclear gene for chloroplast *phytoene synthase* (*Y*1, AY455286.1); *B-transcriptional activator* gene (*b*1, AY078063.2); GASA-like protein gene, (DQ417752.1) and the 22 kDa alpha-*zein* gene cluster (AF090447.2). Chromosome locations of *ZmHack*-2 are labelled: Chr.10 (●), Chr.9 (■), Chr.8 (Δ), Chr.6 (○) and the most invaded chromosome Chr.5 (▼). Scale (0.06) indicates the genetic distances. The NCBI accession numbers are indicated.

2007). Some of these developed from transposase genes of TEs (Cowan et al., 2005) such as *Mariner*/*Pogo*, *hAT*, *piggyBac*, *P*, *Harbinger*, and *Transib* (Table 2). In response, host genomes develop defense mechanisms to control the activity of TEs and their mutagenic potential (Ma et al., 2004; Mansour, 2007, 2008; Blumenstiel, 2011). Relationships between ancestral and domesticated TE-derived eukaryotic genes are evident based on sequence similarity and functional analyses (Figs. 3 and 4). Below we provide a review of the evidence for the different eukaryotic genes that are derived from TEs. Although there have been several reviews of this topic in the past decade (Feschotte et al., 2002; Robertson, 2002; Kazazian, 2004; Gotea and Makalowski, 2006; Volff, 2006; Jurka et al., 2007), our contribution is different in five ways: (1) previous reviews did not have access to a number of recently completed nuclear genome sequences across eukaryotes; (2) earlier reviews did not include comprehensive coverage of the different families of transposable elements across eukaryotes; (3)

Table 1
Functional differences between autonomous and domesticated TEs.

Ancestral autonomous TEs	Domesticated and neofunctionalized TEs
Multi-copy	Single-copy
Present at different positions in the genomes of divergent species	Detectable at orthologous loci in different organisms
Transposition sequences are functional and complete	Transposition sequences are non-functional or missing
Defective TEs inactivated through frameshift mutations	Mutations do not affect the functional open reading frame in different species
Non-synonymous nucleotide substitution	Synonymous nucleotide substitutions

Table 2
Examples of eukaryotic genes originating from transposable elements (TEs), retroviruses and microRNA (MIR) genes during the evolution. # - the section in the text where the gene is discussed.

#	Gene name	Function	Origin	Organism	Reference(s)
2. DNA transposons, genes evolved from transposases of TEs					
2.1.1	*rag*1; *rag*2	V(D)J recombination, a site-specific somatic recombination of the variable region of B cell receptor/immunoglobulin and T cell receptor genes	*Transib* DNA transposase	Jawed vertebrates	Kapitonov and Jurka, 2004, 2005; Volff, 2006
2.1.2	*harbi*1	Involved in DNA rearrangements	*Harbinger* transposase	Fish, birds, frogs, and mammals	Kapitonov and Jurka, 2004
2.2	*haT-like* transposase-derived genes	Involved in reproduction	*Mariner or Pogo* transposase (?)	Mammals, plants, and fungi	Smit and Riggs, 1996; Tudor et al., 1992; de Jesus et al., 2012
2.2.1	*Daysleeper*	Plant Development and regulation of DNA repair gene *Ku*70.	*hAT* transposase	*Arabidopsis thaliana*	Bundock and Hooykaas, 2005
2.2.2	*Gary*	Specific function still unknown	*hAT* transposase	Grasses	Muehlbauer et al., 2006
2.2.3	SchAT	Transposase with DDE motif	*hAT* transposase	Sugar Cane	
2.2.4	*tramp* (*Zbed*1)	Encodes a putative protein similar to Ac-like transposases located at pseudoautosomal region of the human X and Y chromosomes	*hAT* transposase	Mammals (human)	Esposito et al., 1999
2.3	*Pogo* transposase	Centromere-associated protein CENP-B, which binds to the centromeric 17-base-pair Cenp-B box.	*pogo* transposase	*Drosophila*, mammals, plants and fungi	Smit and Riggs, 1996
2.4	*Mariner*-type transposase *Setmar* (*Metnase*)	Increases resistance to ionizing radiation and non-homologous end-joining repair of DNA double-stranded breaks, and promotes integration of exogenous DNA into the host cell genomes	*Mariner*-type transposase	Mouse, human	Robertson and Zumpano, 1997; Lee et al., 2005.
2.5.1	*Mutator*-like transposases *fhy*3; *far*1	Reduced inhibition of hypocotyl elongation, which is specific to Far-Red light and therefore specific to the phytochrome A (*phy*A)-signaling pathway	*Mutator*-like transposase	*Arabidopsis thaliana*	Hudson et al., 2003
2.5.2	*Mustang*	Specific function still unknown	*Mutator* transposase	*Arabidopsis thaliana*, rice, *Medicago truncatula*, poplar	Cowan et al., 2005
2.6.	*P*-elements (*phsa/pgga*)	Encoding 66 kDa repressor-like proteins (RLs) with DNA-binding motif	*P* element transposase	Human, chicken, *Drosophila obscura*, *D. montium*	Pinsker et al., 2001; Reiss et al., 2005
2.7.	*Pgbd*	Specific function still unknown	*PiggyBac*-like transposons	Vertebrates	Sarkar et al., 2003
2.8	*Fob*1p	Regulate the number and the rate of recombination of ribosomal RNA genes	*Maverick*-like element integrase	Yeast (*Saccharomyces cerevisiae*)	Llorens and Marin, 2001; Dlakic, 2002; Dupuy et al., 2011.
3. Retrotransposons-derived eukaryotic genes					
3.1	*Gin*-1	Expressed during embryogenesis and in a various adult human tissues and tumors	*412* and *Mdg1* elements' integrase	Human, mouse, rat, and cow	Llorens and Marin, 2001
3.2. Gag derived genes					
	*PEG*10 *gene* ORF1 and ORF2	Parthenogenetic development and embryonic lethality	*Gypsy*-like LTR retrotransposons	Mammals	Lux et al., 2005; Ono et al., 2001; Volff et al., 2001
	*Rtl*1 or *PEG*11	Paternally expressed genes	*Gypsy*-like GAG and RT	Mammals	Seitz et al., 2003
	MyEF-3	Transcription factor potentially regulating the expression of the myelin basic protein (MBP)	*Gag* proteins	Mouse	Steplewski et al., 1998
	Mart gene family	Embryonic development and control of cell proliferation and apoptosis	*Sushi Ty3/gypsy* LTR retrotransposons	Fish, amphibians, placental mammals	Poulter and Butler, 1998; Brandt et al., 2004; Youngson et al., 2005
	*peg*10	Parthenogenetic development and embryonic lethality	*Gag* proteins retroviral/ retrotransposon (*Ty3/Gypsy*)	Human and mouse	Ono et al., 2001
	Ma gene family	Autoimmune response associated with paraneoplastic neurological disorders	*Gag* proteins retroviral/ retrotransposon (*Ty3/Gypsy*)	Human and mouse	Dalmau et al., 1999; Schüller et al., 2005; Wills et al., 2006
	ma-1/*map*-1	Phosphoprotein expressed, restricted to the brain and testis	*Gag* proteins retroviral/ retrotransposon (*Ty3/Gypsy*)	Human and mouse	Dalmau et al., 1999

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001

Table 2 (*continued*)

#	Gene name	Function	Origin	Organism	Reference(s)
	ma-3	Phosphoprotein expressed, restricted to the brain and testis	*Gag* proteins retroviral/ retrotransposon (*Ty3/Gypsy*)	Human and mouse	Wills et al., 2006
	*Fv*1 (Friend virus susceptibility 1)	Restricts murine leukemia virus replication	*Gag* proteins retroviral/ retrotransposon (*Ty3/Gypsy*)	Human	Best et al., 1996; Benit et al., 1997
3.3. Telomerases with phylogenetic relationships to reverse transcriptases of retrotransposons					
	(a) *TART* telomerase-like activity. (b) *Het*-A telomerase-like activity	(a) RNA-dependent RNA polymerases (RdRpols), telomere maintenance and a target of the RNAi-based system. *TART* transcripts localized in nurse cells. (b) *HeT-A* transcripts accumulate in the oocyte	(a) Non-LTR retrotransposons *A TART*. (b) Non-LTR retrotransposons *Het-A*	*Drosophila*	Eickbush, 1997
	TAHRE telomerase-like activity	Telomere maintenance and a target of the RNAi-based system. *TAHRE* transcripts accumulate in the oocyte	*TAHRE* retrotransposons	*Drosophila*	Shpiz et al., 2007
4. Retrovirus-derived genes (Genomic fossils), Env-derived genes					
	syncytin-1	Fusogenic membrane involved in human placental morphogenesis glycoprotein	*HERV-W* retrovirus	Human, chimpanzee, gorilla, orangutan and gibbon	Mi et al., 2000; Villesen et al., 2004
	syncytin-2	Fusogenic property conserved in simians and has potential role in placenta formation	*HERV-W* retrovirus	Human	Blaise et al., 2003
	syncytin-A; *syncytin*-B	Potential role in placenta formation	Endogenous retrovirus *ERV*	Mouse and other Muridae	Dupressoir et al., 2005
	Iris	Defensive mechanisms against baculoviruses and insect retroviruses	Kanga BEL-like retrovirus	*Drosophila*	Malik and Henikoff, 2005
5. microRNAs, TE-derived miRNA genes					
	mir-136; *mir*-127 micro RNAs	*Rtl*1 RNAi in embryo brain or placenta processed pseudogenes	*MIR* (*SINE*) and *L2* (*LINE*), *5 Alus*, micro RNAs	Mammals, mouse	Seitz et al., 2003; Lin et al., 2007; Smalheiser and Torvik 2005, 2006; Borchert et al., 2006

we demonstrate a correlation between genome sizes and the percentage of transposable elements (TE%); (4) we classify different domesticated transposable elements based on their origin; and (5) we include phylogenetic analyses for selected families of domesticated transposons.

2. DNA transposon-derived eukaryotic genes

2.1. Genes evolved from transposases (tnp) of TEs

There are about 30 eukaryotic genes evolved from DNA transposons (Table 2) (Robertson, 2002) and several of the well-studied examples are reviewed below.

2.1.1. Rag1 and rag2 genes

The enzyme *rag*1 is the most ancient host protein derived from TEs (Kapitonov and Jurka, 2004). *Rag*1 originated almost 500 million years ago (Mya) from a *Transib* DNA transposase, which has a common ancestor in jawed vertebrates. *Rag*1 now encodes a key enzyme with nuclease/transposase-like activities (Kapitonov and Jurka, 2005). Both *rag*1 and *rag*2 initiate V(D)J (Variable, Diverse, and Joining) recombination, a site-specific somatic recombination necessary for the assembly of the gene products of the variable regions of immunoglobulins of B- and T cell receptors. The *rag*1-2 complex can function as a transposase and catalyze intermolecular transpositions *in vitro*. It also functions as an endonuclease, which causes double-stranded breaks in the DNA close to the specific recombination signal sequences (Volff, 2006). The *rag*1-based immune system is the only example of host machinery that evolved from transposases of TIR (*terminal inverted repeats*) DNA transposons (Kapitonov and Jurka, 2005). Several transposase-derived genes show high similarity to DNA/RNA binding proteins (Toth et al., 1995; Jurka and Kapitonov, 1999; Liu et al., 2003).

2.1.2. Harbi1 gene

*Harbi*1, a transposase-coded protein with DNA endonuclease activity, evolved from the transposase gene of *Harbinger* TEs. This gene is present in protists, plants, insects, worms, and vertebrates (Fig. 5). *Harbi*1 encodes two proteins developed from a common ancestor, and subsequently spread in fish, birds, frogs and mammals (Kapitonov and Jurka, 1999, 2004; Volff, 2006). *Harbinger*3_*DR* transposon was identified recently in zebrafish (Sinzelle et al., 2008).

2.2. The hAT-like transposase-derived genes

2.2.1. Daysleeper in Arabidopsis

Genes of transposases of *hAT* (*hobo/Ac/Tam*) TE superfamily were domesticated in various eukaryotes.

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001

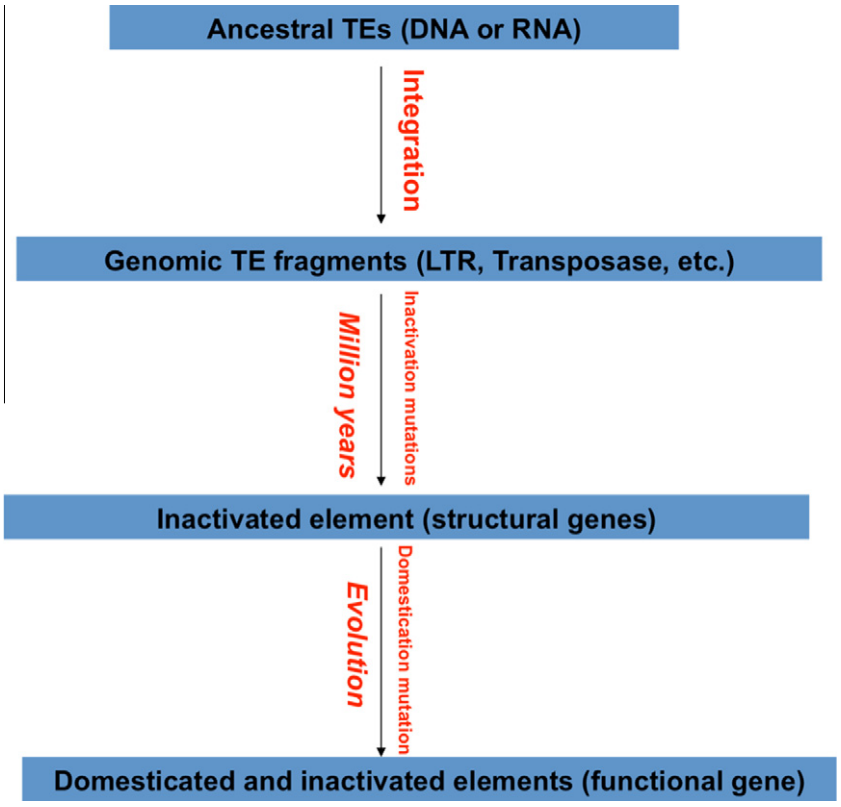

Fig. 3. Molecular events of the domestication of transposable elements to functional eukaryotic genes.

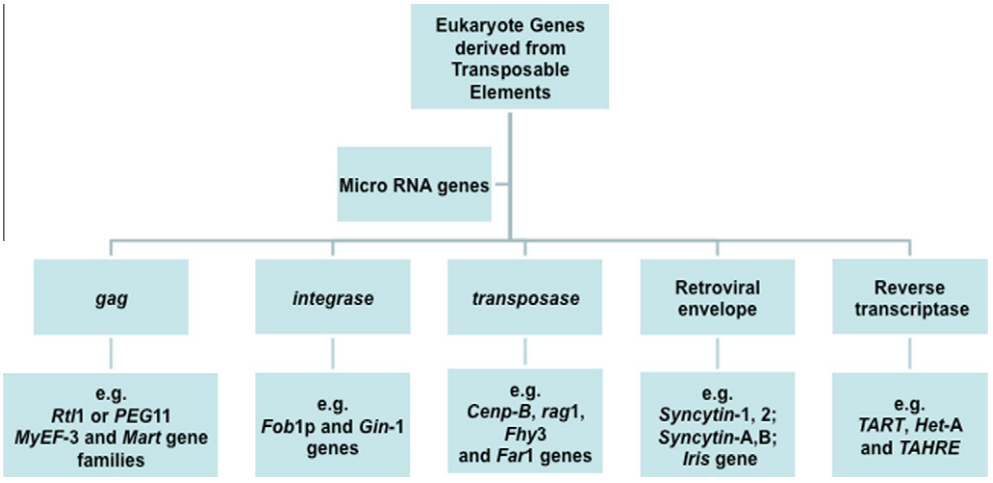

Fig. 4. Examples of different TE genes (e.g. *gag, intergrase* and *transposase*) that were domesticated by host genomes and evolved new cellular functions.

Daysleeper, identified in *Arabidopsis*, encodes a protein that binds to an upstream region of the DNA repair gene *Ku*70 and plays crucial role in plant body development (Bundock and Hooykaas, 2005; Lin et al., 2007).

2.2.2. Gary in cereals

Gary, a member of the *hAT* transposase-like gene family, was domesticated by several cereal genomes at least 60 Mya (Muehlbauer et al., 2006). One or two copies were found in barley, located on the distal end of the long arm of chromosome 2H. Two diverged copies were identified in rice on chromosome 4, and two copies in hexaploid wheat on chromosome 2. However, no homologues were found in *Arabidopsis*. The absence of some key amino acids, required for transposase activity, indicated how TE-derived genes lose transposition activity (Muehlbauer et al., 2006).

2.2.3. SchAT in sugarcane

The hAT transposons are a more heterogeneous group characterized by at least two transposon lineages. Recent

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001

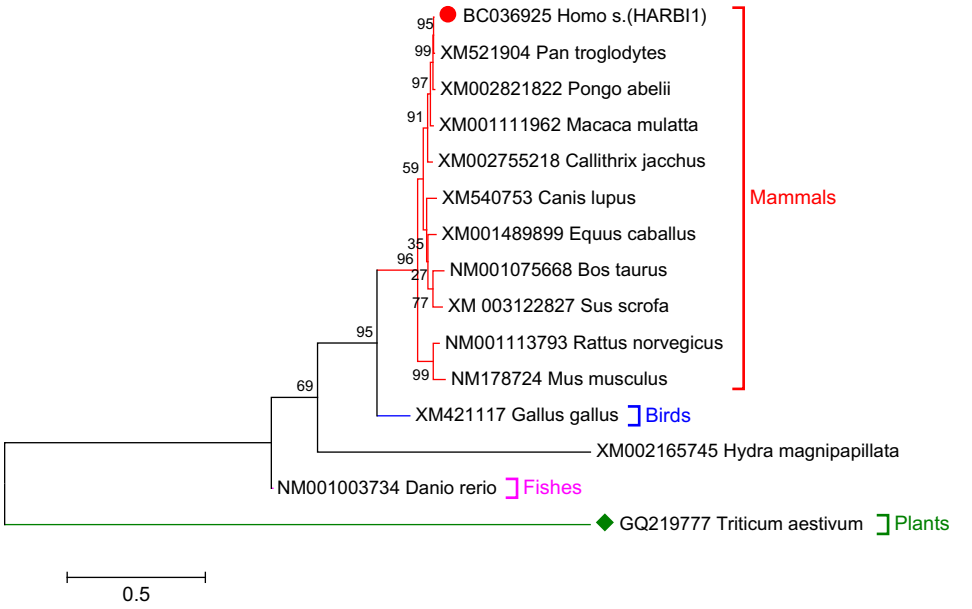

Fig. 5. Bootstrap consensus Maximum Likelihood (Hillis et al., 1994) tree (Mega4, Tamura et al., 2007) derived from the mRNA sequences of the *transposase* genes (aligned sequences of 2101 nt stretches) of *Harbinger* (*harbi*1) DNA transposons (Kapitonov and Jurka, 1999). Bootstrap support values from 1000 replicates are provided at each node. Scale (0.5) indicates the likelihood score. Genbank accession numbers are indicated for each species.

findings in the sugarcane genome showed the existence of at least three different ancestries of *hAT*-like transposase paralogues, including the new domestic transposon SChAT, with a preserved DDE motif, (de Jesus et al., 2012). It is noteworthy that other plant genomes such as wheat, barley and rice, have analogous, expressed transposon-derived genes with no assigned function (Muehlbauer et al., 2006).

2.2.4. Tramp in mammals

The *tramp* gene in mammals encodes a putative protein with an amino acid structure similar to transposases of the *Ac* element of DNA TEs (Esposito et al., 1999). It is located in the Xp/Yp PAR region of a mammals's sex chromosome X, and has a functional homologue on the Y chromosome. It is flanked by putative terminal inverted repeats (TIRs) and has a duplicate target site, which indicates its ancient origin from a transposon. *Tramp* is very divergent in the mouse genome (Gianfrancesco et al., 2001).

2.3. Pogo transposase-derived genes

The *pogo* DNA transposon superfamily was identified in *Drosophila* and later in many other eukaryotic genomes, which suggests a horizontal gene transfer among the species of fungi, nematodes and flies (Casola et al., 2008). One of the genes derived from *pogo* transposase is the centromere-associated protein (*cenp*-B), which was the first sequence-specific DNA–protein complex detected in the centromeric region of human chromosomes (Muro et al., 1992). This protein specifically binds to the 17 base-pair *cenp*-B box of the centromeric region, and was found conserved in mammalian species (Yoda et al., 1992). Gene knockout mutants revealed that *cenp*-B is involved in cellular reproduction rather than in centromere-related functions (Fowler et al., 2000; Tomascik-Cheeseman et al., 2002), and that the gene evolved from transposase genes of TEs (Tudor et al., 1992; Smit and Riggs, 1996). In yeast (*Saccharomyces cerevisiae*), three *cenp*-B homologues were found (Irelan et al., 2001), however, there was a low similarity between yeast and mammalian *cenp*-B (lower than 30%), suggesting that it may have evolved independently in yeast from the mammalian *Mariner/Pogo* transposases (Jurka et al., 2007).

2.4. Mariner-type transposases

Setmar, which encodes the protein metnase, evolved from a *Mariner*-type transposase. The protein fuses to a region of the set domain with histone methyltransferase activity, and increases the resistance to ionizing radiation (Robertson and Zumpano, 1997). *Setmar* also promotes the integration of exogenous DNA into the genome (Lee et al., 2005). The human homologues of ScPSO4/PRP19 (*hPso*4) form a stable complex with metnase of both TIR and non-TIR DNA transposons (Beck et al., 2008).

2.5. Mutator-like transposases

In plants, many genes were domesticated from transposase genes of *Mutator*-like DNA transposons in both monocots and eudicots (Cowan et al., 2005). The Robertson's Mutator DNA TEs in maize undergo cycles of activity and inactivity that correlate with the changes in cytosine methylation. The maize *Mu*4 DNA transposon (NCBI X14224; 2233 bp) includes the mobile element sequence (101 to 2125 bp) and the flanking repeat regions at the 5′ (92 to 100 bp) and 3′ ends (2116 to 2124 bp) (Talbert et al., 1989). Mutator-like elements in *Arabidopsis* are

heavily methylated and inactive (Singer et al., 2001; Cowan et al., 2005), however, in the *ddm*1 (*decrease in DNA methylation*) mutant they became reactive after demethylation (Singer et al., 2001).

2.5.1. Fhy3 and Far1

In plants, *far*1 (gene *far-red impaired response* coding for FAR1 protein, 164 entries available at NCBI, Jan. 2012), *fhy*3 (gene *far-red elongated hypocotyl*) coding for FHY3 protein; 10 entries available), and *frs* (*far*1-*related sequences*), the only transposon-like genes with known host functions, are related to a DNA transposon family of Mutator-like elements (MULEs) (Cowan et al., 2005). *Fhy*3 and *Far*1 (Figs. 6a, b), which are involved in the phytochrome-A signaling pathway, were discovered in *A. thaliana*. There are 12 *Fhy*3/*Far*1-related genes in the *Arabidopsis* genome, and the predicted sizes of encoded proteins range from 531 to 851 amino acids with 12.0% to 82.4% sequence identity (Lin and Wang, 2004). *Far*1 and *Fhy*3 mutants are hyposensitive to far red light. The proteins encoded by *Far*1 and *Fhy*3 genes are related to transposases of type II MuDR family (Hudson et al., 2003), with activities of both DNA-binding and transcriptional activation. These functions indicate that *far*1 and *fhy*3 control the expression of the target genes that bind to the TIRs of DNA transposons (Lin et al., 2007).

2.5.2. Mustang transposase gene family

Mustang is a novel family of domesticated transposase genes of Mutator DNA TEs. These genes are present in both monocots (*Oryza sativa*) and eudicots (*Populus spp.*, *Arabidopsis thaliana* and *Medicago truncatula*), suggesting an ancient domestication event (Cowan et al., 2005).

2.6. Domestication of P-elements

P-transposase encodes a protein with a predicted DNA-binding motif. It was identified in humans and chicken and was domesticated about 300–450 Mya ago. In *Drosophila*, two independent domestications of the *P* transposable element were identified in *D. obscura* and *D. montium*. In *D. montium*, this gene potentially encodes a 66 kDa repressor-like (RL) protein with DNA-binding activity (Pinsker et al., 2001; Reiss et al., 2005).

2.7. Domestication of transposase from piggyBac-like DNA transposons

PiggyBac is a short IR-type (inverted repeat) DNA TE. Many *piggyBac*-like sequences have been indentified in the genomes of fungi, plants, insects, crustaceans, urochordates, amphibians, fishes and mammals (Sarkar et al., 2003). Many other genes were detected in eukaryotic genomes that evolved from *piggyBac*-like transposons, such as *looper* (in humans), *pigibaku* (in *Takifugu rubripes*), *Tx* (in *Xenopus*) and *pokey* (in *Daphnia*). The *piggyBac* family has been useful in biotechnology for genetic transformations (Sarkar et al., 2003).

Pgbd genes found in many vertebrate genomes are also derived from *piggyBac*-like transposons. Five *pgbd* genes

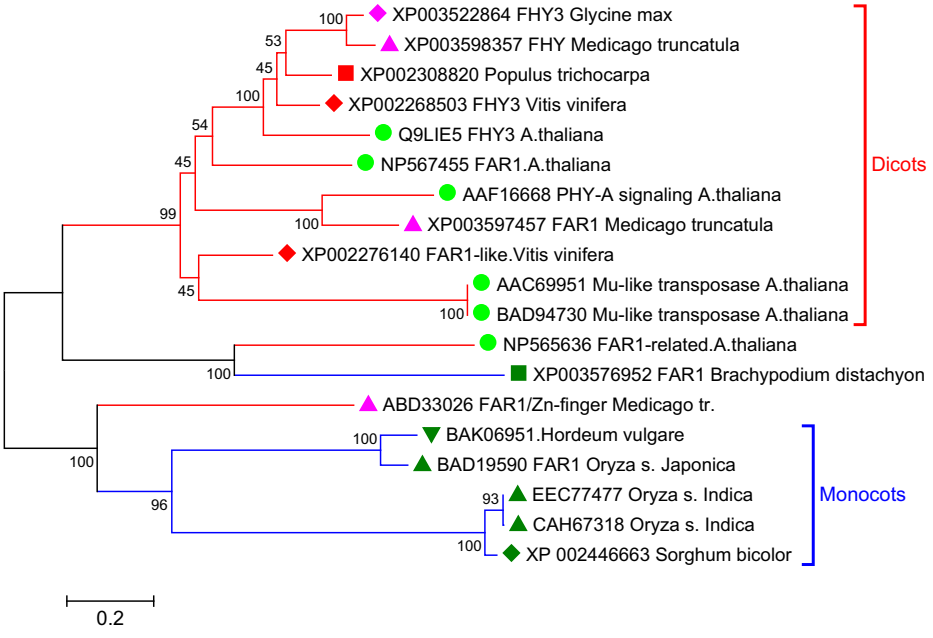

Fig. 6a. Bootstrap consensus Maximum Likelihood (Hillis et al., 1994) tree (Mega4, Tamura et al., 2007) derived from the protein sequences of plant FHY (*far-red elongated hypocotyls*) and FAR1 (*far-red impaired response*) encoded by Mutator-like DNA transposons. The FHY3 sequence (839 aa, *Arabidopsis thaliana*) was downloaded (Swiss-Prot Q9LIE5.1), blasted (NCBI; Altschul et al., 1997) and the consensus stretches of sequences (961 aa, due to deletions) were aligned and tree was computed. Symbols for monocots, eudicots, and woody species are indicated with different colored symbols. Bootstrap support values from 1000 replicates are provided below each node. Scale (0.2) indicates the likelihood score. Genbank accession numbers are indicated for each species.

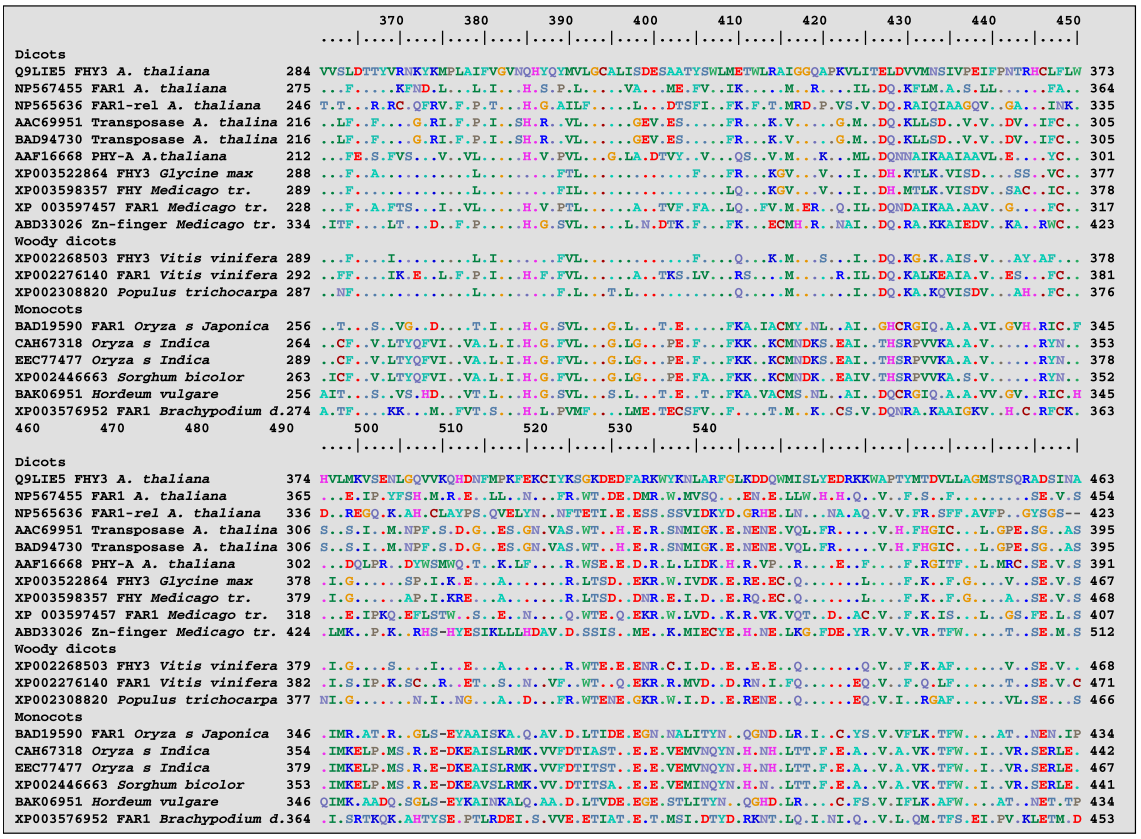

Fig. 6b. Parts of consensus sequence alignments of the highly conserved sequences of plant FHY (*far-red elongated hypocotyls*) and FAR1 (*far-red impaired response*) proteins encoded by *Mutator*-like DNA transposons domesticated during the evolution.

were identified in the human genome and four are present in mouse. *Pgbd5* was also found in the pufferfish (*T. rubripes*), suggesting a domestication event that coincides with the origin of bony vertebrates (Sarkar et al., 2003).

2.8. Fob1p, encoding for transposase/integrase-like protein

In yeast, *fob1p* encodes a blocking protein that regulates the number and rate of ribosomal RNA genes (Dlakic, 2002). Multiple sequence alignment of *fob1p* suggests that this gene is related to c-*integrases* but it does not have the DDE (AspAspGlu) catalytic motif, which facilitates *fob1p*-mediated formation of extrachromosomal DNA circles to accelerate recombination events. The FOB1P protein was probably derived from a transposase/integrase of *Maverick*-like DNA TE element or an LTR retroelement (Llorens and Marin, 2001).

3. Retrotransposon-derived eukaryotic genes

3.1. Genes derived from integrase

Many ancient eukaryotic genes encoding proteins show sequence similarity to DDE-type (AspAspGlu) integrases of retroelements. Such genes have been identified in human, fish and worms (Llorens and Marin, 2001). The difference between the integrase-derived domesticated genes and the ancestral integrase of retrotransposons is the lack of DDE catalytic motif in integrase-derived genes. The non-mammalian integrase may have developed from a giant DNA TE family *Maverick* (Gao and Voytas, 2005).

The sequence of a single copy human gene *gin*-1 (*Gypsy integrase*-1), located on chromosome segment 5q14–5q21, was reconstructed from cDNA sequences (Llorens and Marin, 2001). G*in*-1 encodes an integrase-like protein activated during embryogenesis and tumor development. It has the motifs HHCC (HisHisCysCys), DDE (AspAspGlu) and GPY/F (GlyProTyr/Phe), which are found in many retroviral and retrotransposon integrases. The gene originated from the integrase of *Ty*3/*gypsy* LTR retrotransposon, which is related to the *Mdg*1 elements of *D. melanogaster*. Partial orthologous cDNAs of *gin*-1 were also identified in mouse, rat and cow genomes (Llorens and Marin, 2001).

3.2. Gag-derived genes

Sequencing of the human and mouse genomes revealed over 50 protein-encoding genes that are syntenic and evolved from the *gag* gene of *Gypsy* LTR retrotransposons. The significance of this finding is that *Gypsy* LTRs are the most active LTRs, hence the name for this retrotransposition activity (Llorens and Marin, 2001; Ono et al., 2001;

Volff et al., 2001; Kapitonov et al., 2004; Brandt et al., 2005; Campillos et al., 2006).

*PEG*10 (*paternally expressed gene* 10) is also a *Gypsy* LTR-derived gene, which includes the *gag* gene and a protease domain fused together through ribosomal frame-shift mechanisms (Ono et al., 2001; Volff et al., 2001; Lux et al., 2005). Analysis of two predicted open reading frames (ORFs) revealed that both have homology to *gag* and *pol* genes of retrotransposons of vertebrates. Knockout mice showed that *PEG*10 is important for mouse parthenogenetic development, and thus embryonic lethality (Ono et al., 2006; Jurka et al., 2007).

R*tl*1 (*retrotransposon-like* gene) is the only example of a mammalian gene that evolved from both *gag* and *reverse transcriptase* (RT) of the *Gypsy* retrotransposon (Seitz et al., 2003). More than 50% of all *gag*-derived genes are located on the mammalian X chromosome (Jurka et al., 2007).

MyEF-3 was described as a transcription factor with regulatory activity on the expression of the gene encoding myelin basic protein (MBP) in mouse (Steplewski et al., 1998). This gene may also encode a novel protein in mouse brain that interacts with the MBP MB_1 domain. *MyEF*-3 binds efficiently to double-stranded mb_1 as well as the single-stranded non-coding strand of mb_1. The mb_1 regulatory motif binds to and plays an important role in the transcriptional activation of the MBP promoter in transfection assays. *MyEF*-3 also has an important role in cell type and stage-specific expression of MBP during brain development (Steplewski et al., 1998). Sequence similarity between the ORF of this gene and the *gag* sequence of the LTR retrotransposon suggested that this gene was derived directly from TEs (Ono et al., 2001).

Mart genes isolated from fish and amphibians show high sequence similarity to the *gag* gene of *Sushi Ty3/gypsy* LTR retrotransposons (Brandt et al., 2004). *Mart* genes have experienced purifying selection, suggesting that they are not pseudogenes but rather neofunctionalized retrotransposon genes (Brandt et al., 2004). *Mart* genes contain two long ORFs and the first has homology with retroviral *gag* genes. Phylogenetic analyses showed strong homology to *Sushi-ichi* element of pufferfish (Poulter and Butler, 1998). There are at least ten other *gag*-derived genes in the genomes of placental mammals, but they have lost their ability to retrotranspose autonomously (Brandt et al., 2004). The human genome revealed almost eleven *gag*-derived genes based on sequence similarity, and they were found in other mammalian genomes, including mouse (11 genes), rat, dog, cat and cow (about 12 genes). Almost half of these genes are located on the X chromosome, while the others are located on autosomes (Brandt et al., 2004). These genes were shown to be expressed primarily through two alleles in the embryo and placenta. However, at least six *mart* genes were expressed during mouse embryonic development with ubiquitous expression patterns. Three MART proteins contain the conserved *gag*-specific CCHC (CysCysHisCys) zinc finger motif, suggesting their role in nucleic acid binding (Volff, 2006). Two autosomal *mart* genes showed parental imprinting and paternal expression (Youngson et al., 2005). It was also reported that some *mart* genes are differentially expressed in cancer cells and might be involved in the control of cell proliferation and apoptosis (Youngson et al., 2005).

Gypsy-derived *peg*10 is paternally imprinted and located on chromosome 6 of mouse (Lux et al., 2005). The gene shows sequence similarity with an ancient retroviral/retrotransposon (*Ty3/gypsy*), which was integrated in the human genome in a single copy on chromosome 7q21. *Peg*10 contains two overlapping reading frames, *peg*10-*rf*1 *and peg10-rf*1/2, which encode two proteins. *Peg*10-*rf*1 is a *gag*-derived *mart* gene with an essential role in mammalian development. The PEG10-RF1 protein interacts with the transforming growth factor-b receptor *alk*1 (*activin receptor-like kinase*-1), which is involved in the early embryonic angiogenesis in humans (Lux et al., 2005). The binding of PEG10-ALK is mediated by a 200 amino acid motif. *Peg*10-*rf*1 inhibits *alk*1 as well as *alk*5 signaling. The co-expression of *alk*1 and *peg*10-*rf*1 induces morphological changes of neuronal cells. A knockout study of *peg*10/*mart*2 in mice showed early embryonic lethality due to defects in placenta formation (Ono et al., 2006).

Ma is also a *gag*-derived gene family whose protein products show homology to GAG proteins of retroviral or *ty3/gypsy* LTR-retrotransposons (Schüller et al., 2005; Wills et al., 2006). This family of six genes was identified in human and other mammalian genomes. Three of these genes were located on the X chromosome. *Ma* genes encode neuronal proteins that are the target of the autoimmune response associated with paraneoplastic neurological disorders (Dalmau et al., 1999). By analogy to the pro-apoptotic protein MOAP1, a functional interaction was reported between members of *ma* and *bcl*-2 gene families (Schüller et al., 2005).

*Ma*1 encodes a novel 37 kDa phosphoprotein in the brain and testis identified during the probing of a human cDNA library with anti-*ma* serum (Dalmau et al., 1999). *Ma*1/*map*-1 is able to interact with the pro-apoptotic BAX protein to mediate caspase-dependent apoptosis (Tan et al., 2001). *Ma*1 sequences show high sequence identity to *gag* genes of retroviral and *ty3/gypsy LTR* retrotransposons, suggesting that it evolved from LTR retrotransposons (Tan et al., 2001).

*Ma*3 is the third member of the gene family encoding novel neuronal proteins MA, which are expressed in brain and testes with orthologues in murines (Schüller et al., 2005). Bioinformatic analyses revealed a functional ribosomal frame-shift signal in the human paraneoplastic *ma*3 gene. An RNA pseudo-knot in *ma*3 was important for promoting efficient frame-shifting at the 3′-end of the shift site (Wills et al., 2006). Its exact function is still unknown, although serious neurological effects on *ma*3 ectopic expressions in tumor cells indicate their importance in brain development (Wills et al., 2006).

*Fv*1 (*friend virus susceptibility*-1) is also a *gag*-derived gene located on chromosome 4 of the mouse genome (Best et al., 1996). The evolution of *fv*1 in vertebrates may have taken place against a constant background of retroviral infection. FV1 restricts the replication of murine leukemia virus (M-MuLV; Fig. 7); however, it does not block proviral DNA entry into the cell and its reverse transcription. FV1 also prevents proviral DNA transfer into the nucleus and consequently the integration into the genome and

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001

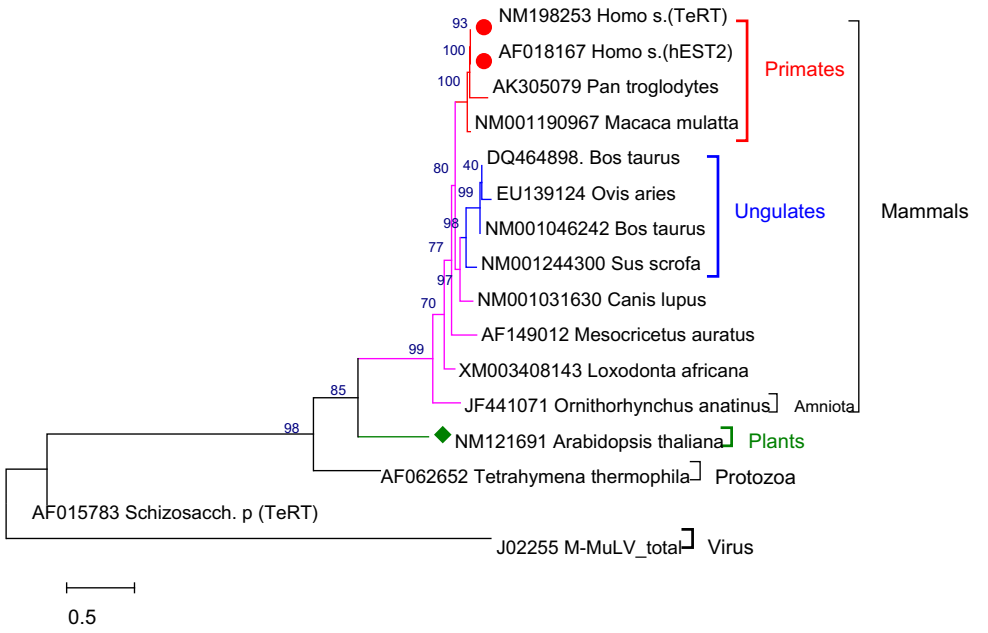

Fig. 7. Bootstrap consensus Maximum Likelihood (Hillis et al., 1994) tree (Mega4, Tamura et al., 2007) derived from the gene sequences TeRT (*telomerase reverse transcriptase*). The sequence *Homos sapiens* (NM198253.2) was downloaded, blasted (NCBI; Altschul et al., 1997) and the consensus stretches of sequences (5009 nt) were aligned and dendrogram was computed. Main clades are indicated, Bootstrap support values from 1000 replicates are provided below each node. Scale (0.5) indicates the likelihood score. Genbank accession numbers are indicated for each species.

formation of provirus, which can prevent or delay viral initiated tumors. FV1 protein and its target may share functional similarities despite the limited nucleotide sequence identity of the encoding genes. It was also reported that FV1 leads to an apparent 50–1000 fold reduction in viral titer *in vitro* (Best et al., 1996).

FV1 protein is highly similar (43% sequence identity) to the GAG protein of some mammalian endogenous retroviruses and quite divergent from the M-MuLV. Interaction between proteins of FV1, GAG, and the viral *gag*-derived capsid protein strongly emphasizes that *fv1* developed from a *gag*-like sequence. Sequence analysis revealed that GAG protein shares almost 43% identity with *fv1* (Benit et al., 1997).

3.3. Telomeres and telomerases

Telomeres comprised of tandem arrays of short repeats are present at the ends of eukaryotic linear chromosomes. During replication telomeres cannot be synthesized by DNA polymerases. Instead, telomerase, a special RNA polymerase, is responsible for telomere replication in most eukaryotes (Greider and Blackburn, 1989). Telomerase is a ribonucleoprotein enzyme that contains a RNA subunit serving as template for telomere replication (Greider and Blackburn, 1989; Nakamura and Cech, 1998). In *D. melanogaster*, telomeres are composed of long arrays of repeated DNA sequences that are produced by successive transposition of two telomere-specific retrotransposons, *HeT-A* and *Tart*. These two elements originated before the separation of the *D. melanogaster* and *D. yakuba* (Casacuberta and Pardue, 2002). *Tahre*, *a* novel element in *Drosophila*, was sequenced by the *Drosophila* Genome Project (Abad et al., 2004). This sequence revealed a conservative element within *HeT-A* and *Tart* that functions in telomere maintenance and serves as a target for the RNAi-based system called the repeat-associated short interfering (rasi) RNA-mediated mechanism (Shpiz et al., 2009).

Sequence similarities between telomerases and non-LTR retrotransposons were noted by Nakamura et al. (1997) and Malik and Eickbush (2001). Because the active centrums of telomerase enzymes share sequence and functional similarity with reverse transcriptases (RTs) they were named telomerase reverse transcriptase (TeRT) (Eickbush, 1997; Fig. 7). The universality of this subunit and the mechanism of telomere addition were found to be similar in *Schizosaccharomyces pombe* and humans. The discovery of non-LTR RTs in bacteria as a form of multicopy single-strand DNA (msDNA), retron (Yamanaka et al., 2002) and retrointrons (Belfort et al., 2002), provided evidence for an ancient origin of non-LTR RTs (Boeke, 2003). It was concluded that the preexisting non-LTR RTs were recruited by eukaryote genomes to perform cellular functions of telomere maintenance (Eickbush, 1997; Pardue and DeBaryshe, 2003). Recently, a connection between telomeres and cell cycle regulation was shown to provide clues to understanding human telomere function (Cenci, 2009).

In *Drosophila*, there is no telomerase, however, telomere length is maintained by transposition of three specialized retroelements (Capkova et al., 2008). The telomeres of *Drosophila* are long tandem arrays of two non-LTR

retrotransposons, *HeT*-A and *Tart*. The close relationship of *Tarts* to the non-LTR RTs was based on functional criteria and RT motif analysis (Pardue and DeBaryshe, 2003). Telomerase replication by reverse transcription strongly resembles target-primed reverse transcription similar to non-LTR retrotransposons after nicking of the target site. Thus, it was proposed that telomerases in *Drosophila* evolved from the domestication of RT genes of non-LTR retrotransposons (Arkhipova et al., 2003). Telomerases and non-LTR retrotransposons also show evolutionary links (Eickbush, 1997; Boeke, 2003).

Sequencing of the *D. melanogaster* genome showed a novel element called TAHRE (*Telomere-Associated and HeT-A-Related Element*) located on the telomeric regions of chromosomes (Abad et al., 2004). TAHRE can attach to the broken end of the terminally deleted chromosomes, which indicates a function in the telomere maintenance in *Drosophila* (Shpiz et al., 2007).

4. Retrovirus-derived eukaryotic genes

Genome wide screening of the human genome for retroviral genes with coding capacity revealed many full length copies of viral envelope (*env*) genes derived from the *env* genes of the endogenous retroviruses (ERVs) (De Parseval and Heidmann, 2005). *Env* encodes the viral envelope glycoprotein, which can bind to cellular receptors and mediate virus entry into the target cell. In the human genome, 59 intact viral polyproteins were detected in scattered positions, and 29 of them originated from gammaretroviral *env* genes (Villesen et al., 2004). Some of these *env* genes play a functional role in the formation of placenta, the nutritional and protective interface between mother and the developing fetus. Transcriptional activation of ERVs in placenta contributes to alternation of the neighboring gene expression (Prudhomme et al., 2005). Phylogenetic analyses of ERVs and present-day retroviruses revealed strong similarities to the transmembrane subunit (TM), suggesting a shared ancestry (Prudhomme et al., 2005).

Syncytin-1, which encodes an ENV-like retroviral envelope protein, is the first example of a retroviral ENV-like protein in humans. This gene originated from a defective provirus HERV-W (*Human Endogenous RetroVirus* W). The SYNCYTIN-1 protein is highly fusogenic to membrane glycoproteins, which are involved in human placental morphogenesis (Mi et al., 2000). *Syncytin*-1 strongly facilitates the connections between the mother and fetus. The expression of *syncytin-1* in a variety of cell lines promotes cell fusion and syncytium formation (Mi et al., 2000). It is possible that this gene is involved in placenta formation because it is expressed in the syncytiotrophoblast layer, which originates from fetal trophoblasts. This function is conserved among mammals, including chimpanzee, gorilla, orangutan and gibbon (Villesen et al., 2004).

Syncytin-2 is another placenta-specific *env*-like gene that was identified after genome-wide analyses of the human genome (Blaise et al., 2003; Villesen et al., 2004). *Syncytin*-2 encodes a protein that also has a fusogenic property conserved in mammals. Like *syncytin*-1, *syncytin*-2 has a potential role in placenta formation (Blaise et al., 2003).

In the mouse genome, two envelope genes of *syncytin*-A and *syncytin*-B were identified in an *in silico* search (Dupressoir et al., 2005). Although both are single copy genes and unrelated to any known murine endogenous retrovirus genes, they are related to *env*-derived genes in Muridae, including mouse, rat, gerbil, vole and hamster. Quantitative RT-PCR analysis indicated that both genes are specifically expressed in the syncytiotrophoblast-containing labyrinthine zone. Both genes have the ability to induce cell-to-cell fusion in different cell lines, which suggests different receptor usage during placenta formation (Dupressoir et al., 2005). Based on this, it was suggested that these genes were introduced into the murine lineage by their retroviral vehicle approximately 20 MYA (Dupressoir et al., 2005).

In *Drosophila*, the envelope-derived gene *iris* was suggested to be domesticated from a BEL-like retrovirus *Kanga* at least 25 MYA (Malik and Henikoff, 2005). It was reported that *iris* and other genes of *env*-origin of *Kanga* retroviruses are homologous to those found in *Baculoviruses* and *Roo* retroviruses of insects. Domestications of both *Kanga* and *Roo* retroviruses were shown in fruit fly and mosquito (Malik and Henikoff, 2005). The IRIS protein retains signal peptide and transmembrane domains. However, it lacks the protease cleavage site that is specific to functional ENV proteins. Hence, it was proposed that *iris* originated as a defensive mechanism against baculoviruses and insect retroviruses. *Iris-A* and *Iris-B* were found in the takahashii–suzukii species groups of *Drosophila*, and originated via a tandem duplication of *iris*. The phylogenetic distribution of *iris* genes was used to resolve the relationship among three closely related sibling species of *Drosophila* (*D. simulans*, *D. sechellia*, and *D. mauritiana*) (Malik and Henikoff, 2005).

5. microRNA-derived genes

MicroRNAs (miRNAs) are an abundant class of RNAs 21–25 nucleotides (nt) long. Depending on the degree of complementarities with their target genes, miRNAs trigger translation by interacting with the target mRNAs (RNA interference, RNAi) (Seitz et al., 2003). The involvement of miRNAs in gene regulation is similar to the antagonistic relationship between TEs and the host genome. Hence, it was suggested that microRNA genes evolved from TEs (Jurka et al., 2007). It was reported that the expression of TEs and the generation of repetitive DNA are coupled with RNA degradation and DNA methylation, which might be mediated by small RNAs (sRNAs) derived from the targeted repetitive DNA (Chan et al., 2005; Vaughn and Martienssen, 2005; Buchon and Vaury, 2006; Qi et al., 2006). Consequently, eukaryotic cells recruit RNAi as a defensive RNA-silencing mechanism to control the replication of viruses and transposable elements (Buchon and Vaury, 2006). There are several lines of evidence concerning the contributions of TEs to the origin and expression of miRNAs involved in gene regulation. In *A. thaliana*, for example, the epigenetic regulation of some endogenous

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001

genes evolved from the silencing mechanisms of TEs (Zilberman and Henikoff, 2005). In addition, it was reported that ancient MIR (SINE) and L2 (LINE) elements were precursors of mammalian microRNAs (Smalheiser and Torvik, 2005).

In the mouse genome, *mir*-136 and *mir*-127 are two miRNA genes located near two CpG islands of *rtl1*, which is a retrotransposon-like gene. *Rtl1* is expressed exclusively from the paternal chromosome, while both *mir*-136 and *mir*-127 are transcribed in an antisense orientation to *rtl1* and expressed from the maternal chromosome. The perfect complementarity between *miR*-136 and *miR*-127 to the ORF of *rtl1* gene suggests that they function as a small interfering RNA to silence *rtl*1 (Hutvagner and Zamore, 2002; Llave et al., 2002) in the placenta and in the embryo brain (Seitz et al., 2003). The *Alu* (SINE) elements and pseudogenes of the human genome might also be the precursors of microRNAs (Devor, 2006; Smalheiser and Torvik, 2006). It was reported that the upstream sequence of *Alu* elements can function as a promoter for the RNA polymerase of miRNAs (Borchert et al., 2006).

6. Conclusions

Molecular evolutionary investigations indicate that DNA transposons were introduced horizontally into the germ-lines of the host genomes by infections millions of years ago (Schaack et al., 2010). However, the idea that transposable elements might flourish because they benefit their host goes back to the discovery of these elements (McClintock, 1984). Some parts of TEs became domesticated with new functions in the cell, while others became inactivated through mutations and by defense mechanisms of the host genome. In both cases, TEs were inherited vertically from generation to generation and inactivated continuously by abiotic and biotic stresses. Most genomes also contain large numbers of retrotransposon sequences that require reverse transcription for their replicative transposition. The endogenous retroviruses (ERVs) likely represent proviral remnants of ancestral germ-line infections that became part of the host genome (Wessler, 1996; De Parseval and Heidmann, 2005; Prudhomme et al., 2005). Theoretically, new cellular function of the domesticated TEs and retroviruses can evolve from any available genes within the cell (Eickbush, 1997). The observations reviewed here support the importance of TEs in genome evolution (Mallet et al., 2004), in the epigenetics of gene expression regulated by microRNAs (Brandt et al., 2005; Devor, 2006; Smalheiser and Torvik, 2006) and their use in biotechnology for genetic transformations (Sarkar et al., 2003).

Acknowledgments

The authors would like to acknowledge the fund of Bibliotheca Alexandrina – Center for Special Studies and Programs, Grant No. 050176. Support was provided by the National Science Foundation to R.K.J. (IOS-1027259). All authors thank two anonymous reviewers for their valuable suggestions.

References

Abad, J.P., de Pablos, B., Osoegawa, K., de Jong, P.J., Gallardo, A.M., Villasante, A., 2004. TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. Mol. Biol. Evol. 21, 1620–1624.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipmand, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Arkhipova, I.R., Pyatkov, K.I., Meselson, M., Evgen'ev, M.B., 2003. Retroelements containing introns in diverse invertebrate taxa. Nat. Genet. 33, 123–124.

Beck, B.D., Park, S.J., Lee, Y.J., Roman, Y., Hromas, R.A., 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.

Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B., Lambowitz, A.M., 2002. Mobile introns: Pathways and proteins. In: Craig, N.L. et al. (Eds.), Mobile DNA II. American Society for Microbiology, Washington, D.C., pp. 761–783.

Belyi, V.A., Levine, A.J., Skalka, A.M., 2010. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: the Parvoviridae and Circoviridae are more than 40 to 50 million years old. J. Virol. 84, 12458–12462.

Benit, L., De Parseval, N., Casella, J.F., Callebaut, I., Cordonnier, A., Heidmann, T., 1997. Cloning of a new murine endogenous retrovirus, *MuERV*-L, with strong similarity to the human HERV-*L* element and with a gag coding sequence closely related to the *Fv1* restriction gene. J. Virol. 71, 5652–5657.

Best, S., Le Tissier, P., Towers, G., Stoye, J.P., 1996. Positional cloning of the mouse retrovirus restriction gene *Fv*1. Nature 382, 826–829.

Biémont, C., Vieira, C., 2006. Junk DNA as an evolutionary force. Nature 443, 521–524.

Blaise, S., de Parseval, N., Benit, L., Heidmann, T., 2003. Genome wide screening for fusogenic human endogenous retrovirus envelopes identifies *syncytin 2*, a gene conserved on primate evolution. Proc. Natl. Acad. Sci. USA 100, 13013–13018.

Blumenstiel, J.P., 2011. Evolutionary dynamics of transposable elements in a small RNA world. Trends Genet. 27, 23–31.

Boeke, J.D., 2003. The unusual phylogenetic distribution of retrotransposons: a hypothesis. Genome Res. 13, 1975–1983.

Borchert, G.M., Lanier, W., Davidson, B.L., 2006. RNA polymerase III transcribes human microRNAs. Nat. Struct. Mol. Biol. 13, 1097–1101.

Brandt, J., Schrauth, S., Veith, A.M., Froschauer, A., Haneke, T., Schultheis, C., Gessler, M., Leimeister, C., Volff, J.N., 2004. Transposable elements as a source of genetic innovation: expression and evolution of a family of retrotransposon-derived neogenes in mammals. Annu. Rev. Genomics Hum. Genet. 345, 101–111.

Brandt, J., Veith, A.M., Volff, J.N., 2005. A family of neofunctionalized Ty3/gypsy retrotransposon genes in mammalian genomes. Cytogenet. Genome Res. 110, 307–317.

Britten, R.J., 2004. Coding sequences of functioning human genes derived entirely from mobile element sequences. Proc. Natl. Acad. Sci. USA 101, 16825–16830.

Britten, R.J., Kohne, D.E., 1968. Repeated sequences in DNA: hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. Science 161, 529–540.

Buchon, N., Vaury, C., 2006. RNAi: a defensive RNA-silencing against viruses and transposable elements. Heredity 96, 195–202.

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

Campillos, M., Doerks, T., Shah, P.K., Bork, P., 2006. Computational characterization of multiple Gag-like human proteins. Trends Genet. 22, 585–589.

Capkova, F.R., Biessmann, H., Mason, J.M., 2008. Regulation of telomere length in *Drosophila*. Cytogenet. Genome Res. 122, 356–364.

Casacuberta, E., Pardue, M.L., 2002. Coevolution of the telomeric retrotransposons across *Drosophila* species. Genetics 161, 1113–1124.

Casola, C., Lawing, A.M., Betran, E., 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.

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

Cenci, G., 2009. *Drosophila* cell cycle under arrest: uncapped telomeres plead guilty. Cell Cycle 8, 990–995.

Chan, S.W., Henderson, I.R., Jacobsen, S.E., 2005. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. Nat. Rev. Genet. 6, 351–360.

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001

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

Curcio, M.J., Derbyshire, K.M., 2003. The outs and ins of transposition: from mu to kangaroo. Nat. Rev. Mol. Cell Biol. 4, 865–877.

Dalmau, J., Gultekin, S.H., Voltz, R., Hoard, R., DesChamps, T., 1999. Ma1, a novel neuron- and testis-specific protein, is recognized by the serum of patients with paraneoplastic neurological disorders. Brain 122, 27–39.

de Jesus, E.M., Cruz, E.A., Cruz, G.M., Van Sluys, M.A., 2012. Diversification of hAT transposase paralogues in the sugarcane genome. Mol. Genet. Genomics 287, 205–219.

De Parseval, N., Heidmann, T., 2005. Human endogenous retroviruses: from infectious elements to human genes. Cytogenet. Genome Res. 110, 318–332.

Debarry, J.D., Ganko, E.W., McCarthy, E.M., McDonald, J.F., 2006. The contribution of LTR retrotransposon sequences to gene evolution in *Mus musculus*. Mol. Biol. Evol. 23, 479–481.

Devor, E.J., 2006. Primate microRNAs miR-220 and miR-492 lie within processed pseudogenes. J. Hered. 97, 186–190.

Dlakic, M., 2002. A model of the replication fork blocking protein Fob1p based on the catalytic core domain of retroviral integrases. Protein Sci. 11, 1274–1277.

Doolittle, W.F., Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284, 601–603.

Dupressoir, A., Marceau, G., Vernochet, C., Benit, L., Kanellopoulos, C., Sapin, V., Heidmann, T., 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proc. Natl. Acad. Sci. USA 102, 725–730.

Dupuy, C., Periquet, G., Serbielle, C., Bézier, A., Louis, F., Drezen, J.M., 2011. Transfer of a chromosomal Maverick to endogenous bracovirus in a parasitoid wasp. Genetica 139, 489–496.

Eickbush, T.H., 1997. Telomerase and retrotransposons: which came first? Science 277, 911–912.

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

Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. 9, 397–405.

Feschotte, C., Jiang, N., Wessler, S.R., 2002. Plant transposable elements: where genetics meets genomics. Nat. Genet. 3, 329–341.

Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., Louis, E.J., 2000. Chromosomal evolution in *Saccharomyces*. Nature 405, 451–454.

Fowler, K.J., Hudson, D.F., Salamonsen, L.A., Edmondson, S.R., Earle, E., Sibson, M.C., Choo, K.H.A., 2000. Uterine dysfunction and genetic modifiers in centromere protein B-deficient mice. Genome Res. 10, 30–41.

Gao, X., Voytas, D.F., 2005. A eukaryotic gene family related to retroelement integrases. Trends Genet. 21, 133–137.

Gianfrancesco, F., Sanges, R., Esposito, T., Tempesta, S., Rao, E., Rappold, G., Archidiacono, N., Graves, A.M.J., Forabosco, A., D'Urso, M., 2001. Differential divergence of three human pseudoautosomal genes and their mouse homologs: implications for sex chromosome evolution. Gen. Res. 11, 2095–2100.

Gotea, V., Makalowski, W., 2006. Do transposable elements really contribute to proteomes? Trends Genet. 22, 260–267.

Greider, C.W., Blackburn, E.H., 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. Nature 337, 331–337.

Hillis, D., Huelsenbeck, J.P., Swofford, D.L., 1994. Hobgoblin of phylogenetics? Nature 369, 363–364.

Hou, Y., Rajagopal, J., Irwin, P.A., Voytas, D.F., 2010. Retrotransposon vectors for gene delivery in plants. Mob. DNA 1, 19.

Hudson, M.E., Lisch, D.R., 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.

Hughes, J.F., Coffin, J.M., 2001. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. Nat. Genet. 29, 487–489.

Hutvagner, G., Zamore, P.D., 2002. A microRNA in a multiple-turnover RNAi enzyme complex. Science 297, 2056–2060.

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

Jurka, J., Kapitonov, V.V., 1999. Sectorial mutagenesis by transposable elements. Genetica 107, 239–248.

Jurka, J., Kapitonov, V.V., Kohany, O., Jurka, M.V., 2007. Repetitive sequences in complex genomes: Structure and evolution. Annu. Rev. Genomics Hum. Genet. 8, 241–259.

Kapitonov, V.V., Jurka, J., 1999. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica 107, 27–37.

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

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

Kapitonov, V.V., Pavlicek, A., Jurka, J., 2004. Anthology of human repetitive DNA. In Meyers, R.A. (Ed.), Encyclopedia of Molecular Cell Biology and Molecular Medicine 1, 251-305, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim.

Kashkush, K., Feldman, M., Levy, A., 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat. Genet. 33, 102–106.

Kazazian Jr., H.H., 2004. Mobile elements: drivers of genome evolution. Science 303, 1626–1632.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

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.

Lin, R., Wang, H., 2004. *Arabidopsis FHY3/FAR1* gene family and distinct roles of its members in light control of *Arabidopsis* development. Plant Physiol. 136, 4010–4022.

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

Lisch, D., Bennetzen, J.L., 2011. Transposable element origins of epigenetic gene regulation. Curr. Opin. Plant Biol. 14, 156–161.

Liu, W., Seto, J., Sibille, E., 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.

Llave, C., Xie, Z., Kasschau, K.D., Carrington, J.C., 2002. Cleavage of scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. Science 297, 2053–2056.

Llorens, C., Marin, I., 2001. A mammalian gene evolved from the integrase domain of an LTR retrotransposon. Mol. Biol. Evol. 18, 1597–1600.

Lux, A., Beil, C., Majety, M., Barron, S., Gallione, C.J., Kuhn, H., Berg, J.N., Kioschis, P., Marchuk, D.A., Hafner, M., 2005. Human retroviral gag- and gag-pol-like proteins interact with the transforming growth factor-beta receptor activin receptor-like kinase 1. J. Biol. Chem. 280, 8482–8493.

Lyon, M.F., 2000. LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? Proc. Natl. Acad. Sci. USA 97, 6248–6249.

Ma, J., Devos, K.M., Bennetzen, J.L., 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Gen. Res. 14, 860–869.

Malik, H.S., Eickbush, H.T., 2001. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Gen. Res. 11, 1187–1197.

Malik, H.S., Henikoff, S., 2005. Positive selection of Iris, a retroviral envelope derived host gene in *Drosophila melanogaster*. PLoS Genet. 1, e44.

Mallet, F., Bouton, O., Prudhomme, S., Cheynet, V., Oriol, G., Bonnaud, B., Lucotte, G., Duret, L., Mandrand, B., 2004. The endogenous retroviral locus *ERVWE1* is a bona fide gene involved in hominoid placental physiology. Proc. Natl. Acad. Sci. USA 101, 1731–1736.

Mansour, A., 2007. Epigenetic activation of genomic retrotransposons. J. Cell Mol. Biol. 6, 99–107.

Mansour, A., 2008. Utilization of genomic retrotransposon as cladistic molecular markers. J. Cell Mol. Biol. 7, 17–28.

McClintock, B., 1984. The significance of responses of the genome to challenge. Science 226, 792–801.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X., Edouard, P., Howes, S., Keith Jr., J.C., McCoy, J.M., 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature 403, 785–789.

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

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., 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.

Nakamura, T.M., Cech, T.R., 1998. Reversing time: origin of telomerase. Cell Cycle 92, 587–590.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W., Lingner, J., Harley, C.B., Cech, T.R., 1997. Telomerase catalytic subunit homologs from fission yeast and human. Science 277, 955–959.

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

Ono, R., Kobayashi, S., Wagatsuma, H., Aisaka, K., Kohda, T., Kaneko-Ishino, T., Ishino, F., 2001. A retrotransposon derived gene, *PEG10*, is a novel imprinted gene located on human chromosome 7q21. Genomics 73, 232–237.

Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., Hino, T., Suzuki-Migishima, R., Ogonuki, N., Miki, H., Kohda, T., Ogura, A., Yokoyama, M., Kaneko-Ishino, T., Ishino, F., 2006. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nat. Genet. 38, 101–106.

Orgel, L.E., Crick, F.H., 1980. Selfish DNA: the ultimate parasite. Nature 284, 604–607.

Pardue, M.L., DeBaryshe, P.G., 2003. Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. Annu. Rev. Genet. 37, 485–511.

Pinsker, W., Haring, E., Hagemann, S., Miller, W.J., 2001. The evolutionary life history of P transposons: from horizontal invaders to domesticated neogenes. Chromosoma 110, 148–158.

Poulter, R., Butler, M., 1998. A retrotransposon family from the pufferfish (fugu) *Fugu rubripes*. Annu. Rev. Genomics Hum. Genet. 215, 241–249.

Prudhomme, S., Bonnaud, B., Mallet, F., 2005. Endogenous retroviruses and animal reproduction. Cytogenet. Genome Res. 110, 353–364.

Qi, Y., He, X., Wang, X.J., Kohany, O., Jurka, J., Hannon, G.J., 2006. Distinct catalytic and noncatalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature 443, 1008–1012.

Reiss, D., Nouaud, D., Ronsseray, S., Anxolabehere, 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.

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

Robertson, H.M., Zumpano, K.L., 1997. Molecular evolution of an ancient mariner transposon, Hsmar1, in the human genome. Annu. Rev. Genomics Hum. Genet. 205, 203–217.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z., Bennetzen, J.L., 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274, 765–768.

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

Schaack, S., Gilbert, C., Feschotte, C., 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol. Evol. 25, 537–546.

Schüller, M., Jenne, D., Voltz, R., 2005. The human PNMA family: novel neuronal proteins implicated in paraneoplastic neurological disease. J. Neuroimmunol. 169, 172–176.

Schulman, A.H., Kalendar, R., 2005. Movable feast: diverse retrotransposons and the contribution to barley (*Hordeum vulgare*) genome dynamics. Cytogenet. Genome Res. 110, 598–605.

Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C., Cavaillé, J., 2003. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat. Genet. 34, 261–262.

Sharp, A.J., Cheng, Z., Eichler, E.E., 2006. Structural variation of the human genome. Nat. Rev. Genet. 7, 85–97.

Shpiz, S., Kwon, D., Uneva, A., Kim, M., Klenov, M., Rozovsky, Y., Georgiev, P., Savitsky, M., Kalmykova, A., 2007. Characterization of *Drosophila* telomeric retroelement TAHRE: Transcription, transpositions, and RNAi-based regulation of expression. Mol. Biol. Evol. 24, 2535–2545.

Shpiz, S., Kwon, D., Rozovsky, Y., Kalmykova, A., 2009. RasiRNA pathway controls antisense expression of *Drosophila* telomeric retrotransposons in the nucleus. Nucleic Acids Res. 37, 268–278.

Singer, T., Yordan, C., Martienssen, R.A., 2001. Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev. 15, 591–602.

Sinzelle, L., Kapitonov, V.V., Grzela, D.P., Jursch, T., Jurka, J., Izsvák, Z., 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.

Smalheiser, N.R., Torvik, V.I., 2005. Mammalian microRNAs derived from genomic repeats. Trends Genet. 21, 322–326.

Smalheiser, N.R., Torvik, V.I., 2006. Alu elements within human mRNAs are probable microRNA targets. Trends Genet. 22, 532–536.

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

Steplewski, A., Krynska, B., Tretiakova, A., Haas, S., Khalili, K., Amini, S., 1998. MyEF-3, a developmentally controlled brain-derived nuclear protein which specifically interacts with myelin basic protein proximal regulatory sequences. Biochem. Biophys. Res. Commun. 243, 295–301.

Talbert, L.E., Patterson, G.I., Chandler, V.L., 1989. Mu transposable elements are structurally diverse and distributed throughout the genus *Zea*. J. Mol. Evol. 29, 28–39.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Tan, K.O., Tan, K.M., Chan, S.L., Yee, K.S., Bevort, M., Ang, C.K., Yu, V.C., 2001. MAP-1, a novel proapoptotic protein containing a BH3-like motif that associates with Bax through its Bcl-2 homology domains. J. Biol. Chem. 276, 2802–2807.

Tomascik-Cheeseman, L., Marchetti, F., Lowe, X., Shamanski, F.L., Nath, J., 2002. CENPB is not critical for meiotic chromosome segregation in male mice. Mutat. Res. 513, 197–203.

Toth, M., Grimsby, J., Buzsaki, G., 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.

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

Vaughn, M.W., Martienssen, R., 2005. It's a small RNA world, after all. Science 309, 1525–1526.

Villesen, P., Aagaard, L., Wiuf, C., Pedersen, F.S., 2004. Identification of endogenous retroviral reading frames in the human genome. Retrovirology 1, 32.

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

Volff, J., Korting, C., Schartl, M., 2001. Ty3/Gypsy retrotransposon fossils in mammalian genomes: Did they evolve into new cellular functions? Mol. Biol. Evol. 18, 266–270.

Wessler, S.R., 1996. Plant retrotransposons: turned on by stress. Curr. Biol. 6, 959–961.

Wills, N.M., Moore, B., Hammer, A., Gesteland, R.F., Atkins, J.F., 2006. A functional_1 ribosomal frameshift signal in the human paraneoplastic *Ma3* gene. J. Biol. Chem. 281, 7082–7088.

Yamanaka, K., Shimamoto, T., Inouye, S., Inouye, M., 2002. Retrons. In: Craig, N.L. et al. (Eds.), Mobile DNA II. American Society for Microbiology, Washington, DC, pp. 784–795.

Yoda, K., Kitagawa, K., Masumoto, H., Muro, Y., Okazaki, T., 1992. A human centromere protein, CENP-B, has a DNA binding domain containing four potential alpha helices at the NH2 terminus, which is separable from dimerizing activity. J. Cell Biol. 119, 1413–1427.

Youngson, N.A., Kocialkowski, S., Peel, N., Ferguson-Smith, A.C., 2005. A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting. J. Mol. Evol. 61, 481–490.

Zilberman, D., Henikoff, S., 2005. Epigenetic inheritance in *Arabidopsis*: selective silence. Curr. Opin. Genet. Dev. 15, 557–562.

Please cite this article in press as: Alzohairy, A.M., et al. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid (2012), http://dx.doi.org/10.1016/j.plasmid.2012.08.001