The role of genes domesticated from LTR retrotransposons and retroviruses in mammals

Tomoko Kaneko-Ishino1* and Fumitoshi Ishino2

1 School of Health Sciences, Tokai University, Isehara, Kanagawa, Japan
2 Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan

*Correspondence: Tomoko Kaneko-Ishino, School of Health Sciences, Tokai University, Bohseidai, Isehara, Kanagawa 259-1133, Japan.
email: tkanekoe@hs.cc.tokai-u.ac.jp

INTRODUCTION

Domestication (exaptation, co-option) is an extended mode of restricting the retrotransposons, endogenous retroviruses (ERVs), and DNA transposons that are integrated into host genomes. It has been proposed that host organisms make use of such transposable DNA elements as a genetic resource of genes for novel purposes (Brosius and Gould, 1992; Stritt, 1999). Telomerase, which maintains the telomere end repeats in chromosomes in eukaryotes, and two recombination activating genes (RAG1 and RAG2) that are essential for producing the vast diversity of immunoglobulin types by V(D)J recombination in vertebrates, are good examples. The former was derived from a reverse transcriptase of a long terminal repeat (LTR) retrotransposon or retrovirus (Nakamura and Cech, 1998) and the latter from a transposase of a DNA transposon (Agrawal et al., 1998; Hiom et al., 1998). Mammalian centromere-associated protein B (CENP-B) facilitates centromere formation and is a DNA-binding protein derived from a transposase of the pogo-like DNA transposon family (Turcan et al., 1999; Kasai et al., 2008). Although it bears considerable similarity to three fusion yeast proteins, ARS-binding protein (Abp1), CENP-B homologs 1 and 2 (Cbh1 and Cbh2), which also exhibit centromere binding, it was recently reported that the origin of mammalian CENP-B is different from that of the three fusion yeast proteins. That is, they are all derived from distinct pogo-like DNA transposons, indicating that convergent domestication occurred in the mammalian and fusion yeast lineages. In plants, the Arabidopsis far-red elongated hypocotyls 3 (FHY3) and far-red-impaired response (FAR1) genes are derived from an ancient Mutator-like transposase, a kind of DNA transposons. They encode transcription factors essential for the light response via phytochrome A signaling (Lisch et al., 2001; Hudson et al., 2003; Lin et al., 2007). From these data, it is clear that the domestication of transposable elements had a profound effect on quite a large numbers of animals and plants during the course of biological evolution, even though only few cases are currently known.

The retrotransposons, ERVs, and their remnant DNA sequences occupy approximately 40% of the mammalian genome and they have long been thought to be either "selfish" genes or useless "junk." Is it thus the case that domestication events are very rare in mammals? Are there any domesticated genes which are present in a mammalian-, therian-, and eutherian-specific manner? Alternatively, are there domesticated genes which have been conserved in a more restricted manner, i.e., species- and strain-specific genes? If so, it would be highly probable that they have contributed to mammalian evolution in various ways and to different degrees.

The recent availability of mammalian genome sequence information enabled us to identify dozens of novel domesticated genes from LTR retrotransposons/ERVs. In 2000, human SYNCYTIN (ERVWE1) was identified as the first candidate domesticated gene derived from ERVs in mammals (Blaise et al., 2003). As shown in Figure 1, it derives from an envelope (Env) gene of a human-specific endogenous retrovirus, HERV-W, and was suggested by in vitro study to mediate placental cytotrophoblast fusion so as to produce syncytiotrophoblast cells in human placental mor-phogenesis. Interestingly, humans have two SYNCYTIN genes, but they are primate-specific genes (Blaise et al., 2003). Similar genes (also called Syncyrins) were also discovered in several mammalian lineages that were independently acquired from Env genes from different ERVs (Dupressoir et al., 2005; Heidmann et al., 2009).
Finally, mouse SyncytinA and B have been to be essential placental genes using knockout mice (Dupressoir et al., 2009, 2011). In 2001, the first and second candidate domesticated genes from a sushi-ichi-related LTR retrotransposon were identified as paternally expressed 10 (PEG10, Ono et al., 2001) and paternally expressed 11/RTL1 (PEG11/RTL1, Chariot et al., 2001). They encode proteins homologous to a retrotransposon Gag and a Pol protein, respectively (Figure 1). Combined with definitive genetic studies using knockout mice, PEG10 and PEG11/RTL1 have been shown to be essential for mammalian development via placenta formation and the subsequent maintenance of its placental function, respectively (Ono et al., 2006; Sekita et al., 2008). As PEG10 is conserved in all the eutherian and marsupial species, it is a therian-specific gene (Suzuki et al., 2007), while PEG11/RTL1 is eutherian-specific (Edwards et al., 2008). All these findings demonstrated that these two domesticated genes are essential in the current mammalian developmental system and indicate that they have been critically involved in the establishment and diversification of viviparous mammals. In other words, these domesticated genes could be major players in the macroevolution of mammals (Kaneko-Ishino and Ishino, 2010).

The concept of macroevolution by such domesticated genes from the LTR retrotransposons/ERVs and the DNA transposons, as well as rewiring gene regulatory networks by non-LTR retrotransposons (Kawahara et al., 2009; Lynch et al., 2011; Schmidt et al., 2012) is a subject of interest not only to biologists, but also to those in the general public who are interested in biological evolution and the origin of human beings. It is of special interest because it implies the existence of a unique long-term relationship between the transposable elements and the emergence of mammals.

In this review, we introduce the essential role played by PEG10 and PEG11/RTL1 in mammalian development via placenta formation, and summarize the current understanding of domesticated genes from the LTR retrotransposons/ERVs, especially those in the mammalian lineages. We also discuss the critically important role of DNA methylation in the process of retrotransposon domestication.

FIGURE 1 | Endogenous retrovirus, LTR retrotransposon and their domesticated genes. Top: an endogenous retrovirus, HVS/LV, and SYNCTYN1. SYNCTYN1 retains LTRs at both ends and the Env gene, while the Gag and Pol genes do not support ORFs corresponding functional proteins because of stop mutations. Bottom: an LTR retrotransposon, sushi-ichi, and the domesticated PEG10 and PEG11/RTL1 genes. Both PEG10 and PEG11/RTL1 have lost LTRs while entire regions exhibit significant homologies to the Gag and Pol genes remaining in some of the retrotransposon domains. The −1 frameshift mechanism is conserved in PEG10, LTR: long terminal repeat, Gag, group-specific antigen, Pol, polymerase, Env, envelope; CCHC, RNA-binding motif; DSG, protease active site; YYDD, reverse transcriptase; DAS, RNase highly conserved motif; HMHC, integrase DNA-binding motif; IDE, strongly conserved integron.

PEG10 AND PEG11/RTL1 IN MAMMALIAN DEVELOPMENT AND EVOLUTION

PEG10 and PEG11/RTL1 were identified as paternally expressed genes in the course of an investigation on genomic imprinting (Charliert et al., 2001; Ono et al., 2001). Genomic imprinting is a mammalian-specific epigenetic mechanism regulating the parent-of-origin expression of a subset of specific genes. For these imprinted genes, the two parental alleles are not equivalent: some of the genes are transcribed only from maternally transmitted alleles (maternally expressed genes, MEGs) and the others are transcribed only from paternally transmitted alleles (paternally expressed genes, PEGs; Kaneko-Ishino et al., 2006). Then, genomic imprinting plays an essential role in mammalian development, growth, and behavior via the activity of certain critically important imprinted genes. In mice, there are more than 10 imprinted regions which have been identified, consisting of both PEGs and MEGs. Among them, a proximal region of chromosome 6 is known to cause early embryonic lethality upon maternal duplication, while maternal duplication of a distal region of chromosome 12 causes late embryonic/neonatal lethality associated with growth retardation (Cattanach and Beechey, 1998; see also Genomic imprinting map: http://www.har.mgu.ac.uk/research/genomic_imprinting/). Mouse Peg10 and Peg11/RTL1 are the major genes responsible for the lethal phenotypes observed in these imprinted regions, respectively (Ono et al., 2006; Sekita et al., 2008). Using knockout mice, we demonstrated that Peg10 and Peg11/RTL1 play essential roles in early placenta formation and maintenance of the placenta in the mid-to-late stages of gestation, respectively. No labyrinth or spongiotrophoblast formation was observed in the placenta of Peg10 knockout mice. The labyrinth layer is a central part of the mouse placenta in which feto-maternal interactions take place. A large portion of the fetal capillaries exist in the labyrinth layer and allow an exchange of nutrients and gases between maternal and fetal blood cells (Figure 2). Mouse embryos require nutrient supply from the placenta starting on day 9.5 of gestation, therefore, Peg10 KO embryos do not survive beyond this stage.

PEG10 encodes two open reading frames exhibiting the highest homology to the Gag and Pol proteins of the sushi-ichi retrotransposon, respectively, and produce two types of proteins, one derived from ORF1 and the other from both ORF1 and 2 (Ono et al., 2001; Voliff et al., 2001; Figure 1). The PEG10 protein retains a CCHC RNA-binding motif in the Gag protein and there is a DSG protease domain in the Pol protein. The −1 frameshift mechanism which produces a Gag-Pol fusion
The genomic record shows that Peg10 is conserved in the eutherian and marsupial mammals among the vertebrates (Suzuki et al., 2007; Kaneko-Ishino and Ishino, 2010). Thus, Peg10 is a very good example of Darwinian evolution and natural selection at work in a macroevolutionary process beyond the individual species which led to the establishment of a subclass of mammals, the therians (Figure 3).

The Peg11/Rtl1 protein also possesses homology to both the Gag and Pol proteins, including the DSG protease domain in the latter, although no frameshift is required in this case (Charlier et al., 2001). The amino acid sequence homology between the Peg10 and Peg11/Rtl1 proteins is approximately 20–30%, indicating their different functions. Mouse Peg11/Rtl1 knockout clearly showed that Peg11/Rtl1 has both a different role from Peg10 and is essential for the maintenance of placental function in the mid-to-late fetal stages (Sekita et al., 2008). Peg11/Rtl1 is expressed in endothelial cells (of extraembryonic mesoderm lineage) of the fetal capillaries in the labyrinth layer, in contrast with Peg10, which is expressed in the labyrinth and spongiotrophoblast cells (of extraembryonic endoderm lineage; Figure 2).

As mentioned above, the fetal capillary is the place where feto-maternal interaction occurs. The loss of Peg11/Rtl1 causes clogging in many of the fetal capillaries in the labyrinth layer because of the phagocytosis of endothelial cells carried out by the surrounding trophoblast cells. The Peg11/Rtl1 protein may protect endothelial cells against placental trophoblast cells, which have a highly invasive and hence dangerous nature, although its biochemical function awaits demonstration. It should be noted that the loss and overexpression of Peg11/Rtl1 are thought to attribute to the etiology of two different human imprint diseases, maternal and paternal disomies of human chromosome 14 (matUPD14 and patUPD14), where Peg11/Rtl1 is located, respectively (Kagami et al., 2008). In these cases, Peg11/Rtl1 plays a major role, not only in placental function, but also in fetal and postnatal growth.

Peg11/Rtl1 is conserved in all eutherian mammals but is absent from marsupial mammals, and is, therefore, a eutherian-specific gene (Edwards et al., 2008). Marsupials use a choriovitelline placenta (yolk sac placenta), which is different from the eutherian choriovallantoic placenta and give birth to their young after a very short gestation period compared to the eutherians (Renfree, 2010). Peg11/Rtl1 function is necessary for the latter to complete their longer gestational period. Therefore, it is probable that Peg11/Rtl1 has a role in the reproduction system of eutherians, which have the choriovallantoic placenta and that it thus contributed to the diversification of these two viviparous mammalian groups. We can say that Peg11/Rtl1 provides another good example of macroevolution in mammals (i.e., the establishment of an infraclass of mammals, the therians) by domesticated genes from LTR retrotransposons (Kaneko-Ishino and Ishino, 2010; Figure 3).

**OTHER SIRH FAMILY GENES DERIVING FROM THE SUSHI-ICHI-RELATED RETROTRANSPOSON**

PEG10 and Peg11/Rtl1 belong to a sushi-ichi-related retrotransposon homolog (SIRH) family consisting of 12 genes, including these two genes as SIRH1 and SIRH2, respectively (Figure 2; Ono et al., 2006). It is also called the mammalian-specific retrotransposon transcripts (MART; Brandt et al., 2005a,b) or SUSHI family (Youngson et al., 2005). The SIRH1–11 genes are conserved in the eutherian species but no marsupial orthologs have been found, yet nevertheless, SIRH2 is derived from a marsupial-specific insertion event (Ono et al., 2011; Figure 3). Among the SIRH family genes, Peg10 (SIRH1), Peg11/Rtl1 (SIRH2), and SIRH9 share homology to both the Gag and Pol proteins, while all the others bear homology only to the Gag protein, but encode proteins of more than 100 amino acid sequences (Brandt et al., 2005a,b; Youngson et al., 2005; Campillos et al., 2006; Ono et al., 2006).
FIGURE 3 | Domestication from LTR retrotransposons and ERVs in mammals. The acquisition of SASPase occurred in a common mammalian ancestor. PEG10 was domesticated in a common therian ancestor while PEG11/RTL1, SIRH3–11, and PNMA1–19 were domesticated in a common eutherian ancestor with subsequent loss of some of the PNMA genes in rodents. The ESCAN domain was domesticated in lower vertebrates and its transition to the SCAN domain took place by combining with the aminoterminal KHAB motifs which had already occurred in certain reptiles. In any event, the expansion of SCAN family is obvious in the eutherians. SIRH12 and PNMA-MS1 are derived from marsupial-specific domestication events. TheSYNCYTins were independently recruited in several different mammalian lineages. The eutherians and marsupials are oviparous, having chorioallantoic and choriovitelline placentas (yolk sac placentas), respectively. Both PEG10 and PEG11/RTL1 are essential for the proper formation of efficient chorioallantoic placentas.

in the corresponding region between the ectodermal-neural cortex (ENC1) and rho-guanine nucleotide exchange factor (RGNEF) genes where wallaby SIRH12 and opossum pseudo SIRH12 are located in the marsupial genome. These findings demonstrate that the SIRH family of genes appeared mainly around the establishment of the therian mammals, one (PEG10) before and all the others after the split of the marsupials and eutherians (Figure 3). Consequently, the eutherians and the marsupials have a different set of SIRH family genes except for PEG10. It is likely that SIRH3–11 and SIRH12 also have roles in the eutherian and marsupial developmental and reproductive systems as well as PEG10 and PEG11. As they are expressed in the brain, ovary, and testis as well as the placenta, they may be related to ovulation, gestation, delivery, and/or maternal nursing behaviors, including lactation, as well as placenta formation. Their respective functions are now under investigation using knockout mice.

PNMA-FAMILY GENES FROM THE Gypsy12_DR-RELATED LTR RETROTRANSPOSON

A paraneoplastic Ma antigen (PNMA) family is another large family consisting of eutherian- and marsupial-specific genes (Schüller et al., 2005; Iwasaki et al., in preparation; Figure 3). PNMA-family genes (PNMA1–3) were first identified as genes encoding neuronal auto-antigens using sera from human patients with paraneoplastic neurological syndromes (Volz et al., 1999; Rosenfeld et al., 2001). By comprehensive search of a protein database, Schüller et al. (2005) identified additional human PNMA genes, MMAP1/PNMA4, PNMA5, and PNMA6, among which PNMA6 has no mouse ortholog. Campillos et al. (2006) performed a genome-wide search for PNMA genes and identified a total 15 genes and 1 pseudogene in humans. They also showed that all of the PNMA genes were related to a Gypsy12_DR-related Gag protein group of the Ty3/Gypsy LTR retrotransposons isolated from zebrafish (Danio rerio) and that there was no Gypsy12_DR-derived
sequences in birds. Recently, Iwasaki and colleagues identified novel PNMA genes by a re-examination of the entire mouse and human entire genome sequences and the PNMA-family genes found thus far number 15 and 19 in mice and humans, respectively; all of these genes have Gag-like proteins, but none are homologous to the Pol protein. The difference in number may be due to the rodent-specific deletion of the PNMA6-C4D genes on X chromosome (Schuller et al., 2005; Iwasaki et al., in preparation).

No knockout mouse studies on PNMA genes have been reported, but there are reports indicating that these genes are involved in important biological pathways and related to human diseases. PNMA4-deficient cells exhibit aggressive anchorage-independent growth, suggesting that PNMA4 has an important role in regulating apoptotic signaling in a strict temporal manner in mammalian cells, because the PNMA4/MOPAP-1 protein is short-lived and constitutively degraded by the ubiquitin–proteasome system (Lee et al., 2009).

Cho et al. (2008a, 2011) reported PNMA10 to be a candidate gene for X-linked mental retardation (XLMR) in humans. In mice, Pnma10(Zcchc12) is expressed in the embryonic ventral forebrain in a cholinergic-neuron-specific manner (Cho et al., 2011), and is known to act as a transcriptional co-activator for bone morphogenetic protein (BMP) signaling by binding to the SMAD family protein (BMP) related to the transcription of IGFBP (insulin-like growth factor-binding proteins) genes and have been identified as two of the genes involved in 3-M syndrome (Huber et al., 2005; Hanson et al., 2009). Importantly, the Ossbl1 protein interacts with both the PNMA14/CDC28 and CDC2 proteins, indicating that this protein complex is responsible for the growth retardation observed in 3-M patients. These findings suggest that PNMA genes play an important role in the embryonic development and growth that are impacted by human diseases.

We have recently identified two marsupial-specific PNMA genes, PNMA-MS1 and -MS2 (Iwasaki et al., in preparation; Figure 3). PNMA-MS2 exists in the same genome location in both the Australian (tammar wallaby) and South American (gray short-tailed opossum) marsupial species, but no orthologs exist in the eutherians, suggesting that they are derived from a marsupial-specific domestication event similar to that of SIRH12. PNMA-MS2 was only found in the opossum because there is a gap in the corresponding region of the wallaby genome sequence. However, it is also clear that PNMA-family genes were independently domesticated in the eutherian and the marsupial lineages, and may have certain eutherian- and marsupial-specific functions, respectively.

THE RETROVIRAL-LIKE ASPARTIC PROTEASE SASPase

Skin aspartic protease (SASPase), which is known a retroviral-like aspartic protease (Bernard et al., 2005), plays a key role in determining the texture of skin by modulating the degree of hydration via the processing of profilaggrin (Matsui et al., 2006, 2011; Barker et al., 2007). SASPase is a single gene conserved in the eutherians, marsupials, and presumably the monotremes (Matsui, personal communication), and thus is a mammalian-specific gene (Figure 3). The profilaggrin protein comprises a tandem array of filaggrin monomers and the SASPase is its specific protease which produces the filaggrin monomer (Matsui et al., 2011). The Filaggrin gene has recently been identified to be etiologically responsible for atopic dermatitis (Barker et al., 2007). Interestingly, both SASPase and profilaggrin are unique to mammals and expressed exclusively in the stratified epidermis in skin. Therefore, it is highly likely that they contributed to the establishment of the mammalian-specific skin barrier system. Aberrant SASPase expression in transgenic mice reportedly leads to impaired skin regeneration and skin remodeling after cutaneous injury or chemically induced hyperplasia (Håkansson et al., 2010), and SASPase-deficient mice exhibit fine wrinkles on the sides of the adult body (Matsui et al., 2006).

SCAN-FAMILY GENES RAPIDLY EXPANDED IN THE COURSE OF EUThERIAN EVOLUTION

The SCAN-family is not a mammalian-specific gene family because its ancestral form exists in non-mammalian vertebrates, but nevertheless, an enormous expansion occurred in the eutherian species (Figure 3). The SCAN motif consists of only a C-terminal portion of the Gag capsid (CA) protein and, in mammals, it always accompanied by multiple C2H2 zinc finger motifs and/or Kruppel-associated box (KRAB) domains neither of which is of retrotransposon origin. It is suggested that the former part was already domesticated at or near the root of the tetrapod animal branch from a full-length CA gene derived from a Gag-like retrotransposon. This is called the extended SCAN (ESCAN) domain and that either it or its truncated SCAN motif combined with the zinc finger and/or KRAB motifs in the Anolis lizard (Emerson and Thomas, 2011). Approximately, 60 and 40 genes are known in humans and mice, respectively, and some of them are involved in development and differentiation as transcription factors, such as ZNF202, ZNF197, ZNF444, ZNF274 (neurotrophin receptor interacting factor, NRIIF), Zfp296 (NSD1-interacting zinc finger protein 1, Nip1p) and Zfp183 (N2E; Edelstein and Collins, 2005).

Therefore, it is highly likely that some of the SCAN-family genes are related to certain eutherian-specific functions. One example is paternal expressed gene 3 (PEG3; Kuroiwa et al., 1996) that was reported to be essential for maternal nursing behavior as well as promoting embryonic growth (Li et al., 1999). The PEG3 protein has very unique structural features among C2H2 zinc finger proteins, such as amino acid sequences for 11 C2H2 zinc finger motifs and a wider spacing of these motifs. The C2H2 zinc finger proteins comprise the largest class of eukaryotic transcription factors, yet no other C2H2 zinc finger proteins have such features (Kuroiwa et al., 1996). PEG3 is widely expressed during fetal development of mice, and strongly in adult neurons and skeletal muscle. The Peg3 KO offspring are approximately 20% smaller at birth, with markedly reduced nursing behavior and a reduced number of oxytocin-positive neurons in the hypothalamus of Peg3 KO females (Li et al., 1999). Human PEG3 has...
tumor-suppressing activity in glioma cell lines by its capacity to inhibit Wnt signaling, and the loss of its expression is reported- 
edly observed in gliomas (Kohda et al., 2001; Mackawa et al., 2004; Jiang et al., 2010).

INDEPENDENT DOMESTICATION EVENTS OF THE SYNCYTINs IN DIFFERENT LINEAGES IN EUHERIANS

As mentioned in Section “Introduction,” SYNCYTIN was first dis- covered in humans (Böndö et al., 2000; Mi et al., 2000). Although there are many Env-related DNA sequences in the human genome, only two exhibit fusogenic activity in cell fusion assays and now these are called SYNCYTIN1 and 2 (Blaise et al., 2003). They are derived from different human-specific ERVs, HERV-W, and HERV-FRD, and became integrated into a primate lineage 23 and >40 MYA, respectively (Figure 3). Recent studies demon- strated that similar genes exist in an order- or family-specific manner in several mammalian lineages, i.e., producing syncytio- tosphoblast cells by cell fusion in the placenta. Mice also have two Syncytin genes, SyncytinA and B, derived from Muri- dae family-specific integrations of HERV-FH-related ERVs(s) approximately 20 MYA (Dupressoir et al., 2005; Figure 3), and rabbits (Oryctolagus cuniculus) have another SYNCYTIN-Ory1 from Leporidae family-specific integration of a different type- D retrovirus 12–30 MYA (Heidmann et al., 2009). Therefore, at least three independent domestication events have been con- firmed in the eutherians, indicating that domestication from ERV’s which were actively functioning during the time of mammalian radiation. Synctyn1 knockout mice exhibit mid-fetal lethality because of the structural abnormality of the placenta (Dupressoir et al., 2009); and Fv1 allele knockout of both Synctytin and B causes an even more severe phenotype, early embryonic lethality (Dupre- soir et al., 2011). Among the eutherians, placental morphology and functions are quite substantially diverged. Therefore, it is very interesting that the SYNCYTINs from the ERVs appear to have important roles in the placenta that they play in an order- or family-specific manner, while PEG10 and PEG11/RTL1 from the LTR retrotransposons are conserved in the therians and eutherians, respectively, and presumably have contributed to the establish- ment of the basic structure of viviparous reproductive systems in the current eutherian species.

RESISTANCE TO VIRAL INFECTION BY DOMESTICATED VIRAL GENES

ERVs have long been thought to confer resistance to infection by enzootic retroviruses. Well-known examples are Friend virus susceptibility 1 and 4 (Fv1 and Fv4) and resistance to mink cell focus-forming (MCF) virus (Rom) genes, which exhibit resistance to murine leukemia viruses (MuLVs) in mice (Pincus et al., 1971; Suzuki, 1975; Hartley et al., 1985). Fv1 is derived from the Gag region of an ancient MERV-L element (Best et al., 1996; Benit et al., 1997), whereas Fv4 and Rom consist of intact Env genes, the expression of which prevents infection via receptor interference (Ikeda et al., 1985; Izu and Kozak, 1996; Taylor et al., 2001; Jiang et al., 2002). Endogenous betaretroviruses (enSRVs) in sheep are another example (Dunlap et al., 2003). The ovine genome pos- sesses approximately 20 copies of enSRVs that are highly related to two enzootic oncoviruses, Jaagsiekte sheep retrovirus (JSRV) and enzootic nasal tumor virus. It has been proposed that the enSRVs Env genes are beneficial to the host and help protect of the uterus from viral infection and act as regulators of placental morphogenesis and function. They exist as species- or strain-specific genes, meaning that they are derived from recent domestication events compared to the SIRH-, PNMA-, and SCAN-family genes as well as the SASPase gene.

Therefore, it is clear that the domestication fromLTR retro- transposons and ERVs has a very long history, dating from around the time of the establishment of vertebrates, on through the estab- lishment and diversification of mammals and ultimately to the radiation of each mammalian species. Koala retrovirus (KoRV) has recently been reported to cause leukemia, lymphoma, and immunosuppression in the Australian Koala population (Tarlinton et al., 2006; Storey, 2006). Interestingly, KoRV is currently undergoing endogenization and it is likely that it entered the koala genome within the last 200 years. Therefore, retrotrans- poson endogenization may be a fairly ordinary process in the long course of evolution, and novel genes may continue to appear by this mechanism in the future.

GENE DUPLICATION OF DOMESTICATED GENES

Although dozens of domesticated genes have been found in mam- mals, this does not necessarily mean that independent domestica- tion events have happened as often as the number of domesticated genes. Certain domesticated genes have apparently been pro- duced by the gene duplication of an originally domesticated gene, such as in the SCAN family of genes. The SCAN domain was domesticated long before the emergence of mammals in the lower vertebrates (ESCAN) and then a new combination of this domain and zinc finger and/or KRAB motifs produced the SCAN- family gene prototype in a reptile, and its expansion occurred during radiation of the eutherians (Emerson and Thomas, 2011; Figure 3).

SIRH4, 5, and 6 as well as PNMA6A, 6B, 6C, and 6D, are other clear examples of gene duplication. The domestication of the orig- inal gene must have occurred in the ancestral eutherian mammals, but these clusters were produced by gene duplication because they encode very nearly the same coding frames. It is interesting to elu- cidate whether they are in the process of diversifying into genes with different functions or there is some as yet unknown reason for them to multiply and increase their copy numbers in this way. Nevertheless, as discussed above, at least two independent domes- tication events occurred in the eutherians and the marsupials in the SIRH and PNMA families, and at least four independent domestici- ations have been confirmed in three different eutherian lineages in the case of the SYNCYTINs (Figure 3).

THE ESSENTIAL ROLE OF DNA METHYLATION IN THE DOMESTICATION PROCESS

Retrotransposons are potentially harmful to host organisms because their integration not only causes genetic diseases by dis- rupting essential genes, but also induces chromosomal deletion as well as recombination by DNA homologous recombination between the two of them. Their integration could also disturb transcription of neighboring genes. Thus, host organisms must
sequences in the host genome? How are they critically silenced and yet stably inherited from generation to generation in a manner similar to endogenous DNA methylation proposed by Kimura (1968, 1983), such as in the mammalian genome. According to the neutral theory of molecular evolution proposed by Kimura (1968, 1983), such neutral mutations are fixed in a population by the mechanism of random drift. Ohta (2002) proposed in her “neutral neutral theory” as an extension of neutral evolution that less harmful mutations could become fixed in a population if the population size were sufficiently small. We previously proposed the hypothesis that in the course of retrotransposon domestication the neutral or nearly neutral evolution preceded Darwinian evolution and helped supply novel genes for novel purposes from the integrated retrotransposons (Kaneko-Ishino et al., 2006; Kaneko-Ishino and Ishino, 2010). In brief, we assume that either neutral or nearly neutral evolution played essential background roles by both inactivating and neutralizing integrated retrotransposons. Subsequently, their gradual conversion from silenced harmful genes to slightly advantageous genes took place as the result of multiple mutations. A loss of such silencing, at least in a subset of tissues, was ultimately required for the “new gene” to have a certain function. Darwinian forces then came into play, and by natural selection certain of these genes became more useful functionally and thus advantageous for the host organisms. It should be noted that in extraembryonic organs, such as the yolk sac and placenta in mammals, the DNA methylation levels are lower than those in other embryonic and adult tissues. Therefore, a leaky expression of retrotransposons and retroviruses constantly occurs. In this situation, the integrated retrotransposons and their subsequent mutated forms would be less harmful. However, in the case of advantageous mutations, a swift transition from the state of nearly neutral evolution to that of Darwinian evolution would take place. In this regard, the extraembryonic tissues might have been a site of retrotransposon domestication during the course of mammalian evolution, which is consistent with the fact that the domesticated PEG10/RTL1 and SYNTYTIN genes play essential roles in the placenta (Kaneko-Ishino and Ishino, 2010).

In this hypothetical scenario, various epigenetic mechanisms, such as DNA methylation and histone modifications, might have played a critical role. In mammals, DNMT1 is the essential maintenance DNA methyltransferase and the loss of its activity causes early embryonic lethality associated with overexpression of IAP retrovirus (Li et al., 1992; Walsh et al., 1998). The two de novo DNA methyltransferase DNMT3A and DNMT3B are also essential for mammalian development, and the loss of their activities causes lethality in the postnatal and embryonic period in mice, respectively (Okano et al., 1999). Overexpression of IAP retrovirus was also observed in Dmnt1 and Dmnt3b double knockout mice, although to a lesser degree than Dmnt1 KO mice (Okano et al., 1999). DNMT3L does not have DNA methyltransferase activity itself, but has an essential function of producing a different DNA methylation status in female- and male-derived genomic DNA in the process of establishing the genomic imprinting memories associated with DNMT3A (Bourc’his et al., 2001; Hata et al., 2002). It is known that this complex is also essential for retrotransposon methylation in the paternal germ line (Bourc’his et al., 2001; Bourc’his and Bestor, 2004). The coincident emergence of DNMT3L in the therian mammals is highly suggestive, both for the origin of the genomic imprinting mechanism as well as the abundance of ITR retrotransposons/ERVs, each of which is specific to the therian genome (Yokomine et al., 2006). It should be noted that H3K9 methyltransferase ERG-associated protein with SET domain (SETD1B) coupled with KRAB-associated protein 1 (KAP1, also called TRIM28) and zinc-finger protein ZFP806 is required for H3K9 trimethylation as well as the repression of the retrotransposons and ERVs in undifferentiated mouse ES cells (Wolf and Goff, 2007, 2009; Matsui et al., 2010). Such a DNA methylation-independent pathway may be necessary, because DNA methylation is dynamically reprogrammed during the early embryonic period in mammals.

Finally, we would like to consider how the mammalian viviparous reproductive system originally started using the retrotransposon-derived PEG10 gene. If this new reproductive system first happened in a single individual, was it possible for such an individual to survive and propagate his or her offspring? Is it worth mentioning that the nearly neutral theory of molecular evolution can also explain how new species originated not from a single individual, but rather from a population subset (Kimura, 1983). Preadaptive mutations were already distributed in a neutral manner. Adaptive functions emerged under the selective pressures of a new environment. This suggests the neutral evolution process could also play a role as an “evolutionary capacitor,” as predicted in the case of heat-shock protein (Hsp) 90, where genotypic variations in other genes are masked and therefore are accumulated without causing any evident phenotypic changes in the chaperone activity of Hsp90 per se (Rutherford and Lindquist, 1998; Bergman and Siegel, 2003). However, this original scenario has recently come under challenge because Hsp90 also acts as a suppressor of retrotransposons and its mutation induces retrotransposon transposition, thus causing a number of secondary mutations (Specchia et al., 2010; Gangaraju et al., 2011).

DNA methylation is commonly observed in a wide range of organisms, from bacteria to plants and animals, although certain model organisms do in fact lack this feature. We propose that changes in DNA methylation in genome regulation systems gave rise to the great diversity of the organisms across the earth. In particular, as mammals developed their particularly specialized DNA methylation system, mammalian evolution was advanced by a series of retrotransposon domestication events. Retrotransposons serve as a double-edged sword in development and evolution, i.e., either harmful or beneficial depending on which time scale is used. The domestication of retrotransposons seems likely to be a very rare event, but once it has taken place, its impact is profound, which is especially the case in mammalian evolution. That may provide the raison d’être for the ITR retrotransposons in the mammalian genome.
ACKNOWLEDGMENTS

We thank all the collaborators and all the laboratory members, especially, Bryska Otos, Yoshiaki Nakano, Shinya Shimotakahara, and Takashi Kohda for analyzing Peg10 and Peg11 KO mice and murine Peg10 and Marilyn Renfrew and Jennifer Graves for several murine research projects. The work has long been supported by a number of grants, Grants-in-Aid for Scientific Research (S) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan, Funding Program for Next Generation World-Leading Researchers (NEXT Program) to Tomoko Kaneko-Ishino, Bilateral Program on Joint Research Project (to Fumitoshi Ishino), and Creative Science Research (to Fumitoshi Ishino and Tomoko Kaneko-Ishino) from the Japan Society for the Promotion of Science (JSPS), Asahi Glass Foundation (to Tomoko Kaneko-Ishino) and Mitsubishi foundation (to Fumitoshi Ishino). Pacific Edit reviewed the manuscript prior to submission.

REFERENCES

Agrawal, A., Eastman, Q. M., and Schultz, D. G. (1998). Transposition mediated by BAG7 and MAR2 and its implications for the evolution of the immune system. Nature 394, 744–751.

Barder, J. N., Dalmann, C. N., Yao, Z., Liu, H., Hou, B. P., Lou, S. F., Allen, M. H., Maggett, S. J., Reynolds, N. J., Trounson, R. C., and McClain, W. H. (2007). Null mutations in the flagerlin gene (flgA) determine major susceptibility to early-onset atopic dermatitis that persists into adulthood. J. Invest. Dermatol. 127, 564–567.

Bent, L., De Pauw, N., Casella, J. F., Callebaut, I., Cordoumy, A., and Heidmann, T. (1997). Cloning of a new murine endogenous retrovirus, MERV-L, with strong similarity to the human HERV-L element and with a gag-pgag coding sequence closely related to the Fv1 restriction gene. J. Virol. 71, 3652–3657.

Bengtson, A., and Sengel, M. L. (2003). Evolutionary capacity as a general feature of complex gene networks. Trends Genet. 19, 424–430.

Bernard, D., Méhul, B., Thomas-Bergman, A., and Siegel, M. L. (2003). Serotypes of pseudoviruses and other “junk DNA” genes is required for formation of the co-opted retroviral envelope syncytin receptor. J. Virol. 74, 3521–3529.

BRouch, D., and Boster, T. H. (2004). Transposition silencing and imprint establishment in mammalian germ cells. Cell 119, 381–388.

Bross, D., Xu, G. L., Lin, C. S., Boll, B., and Boster, T. H. (2005). Dnmt3l and the establishment of maternal genomic imprinting. Sci. 294, 2338–2340.

Brouillette, S., and Voit, J. N. (2004a). A family of nondecalinyl tryptophanostamine genes in mammalian genomes. Cytogenet Genome Res. 107, 307–317.

Brosh, J., Schneid, S., Voit, A. M., Frohnhauer, A., Hanske, T., Schüblin, C., Goebel, M., Leimont, C., and Voit, J. N. (2005b). Transposable elements as a source of genetic innovation: expression and evolution of a family of imprinted retrotransposon-derived genes in mammals. Gene 365, 101–117.

Brosh, J., and Gaul, S. J. (1982). On “germacentric”: a comprehensive (and respectful) taxonomy for pseudogenes and other “junk DNA.” Proc. Natl. Acad. Sci. U.S.A. 89, 10790–10795.

Campillos, M., Doerks, T., Shah, P. K., Bork, P. (2006). Computational characterization of multiple Gap-like human proteins. Bioinformatics 22, 339–350.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.

Casola, C., Hakas, D., and Forteche, P. (2000). “Pervaginale” a comprehensive and respectful taxonomy for pseudogenes and other “junk DNA”. Proc. Natl. Acad. Sci. U.S.A. 97, 10790–10795.
Hildenbrand, M., Blümmel, V., Hartmann, B., Lüdemann, B., Grube, N., Angel, P., and Hoss, J. (2010). Improved skin regeneration and reepithelialization after cutaneous injury and chemically induced hyperplasia in transgenic mice. J. Invest. Dermatol. 130, 1322–1330.

Hoff, K., Murl, M., and Gallert, M. (1998). DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. G3 94, 463–470.

Huber, C., Diao-Suattageta, D., Glaess, A. O., Sullivan, J., Brauer, R., Wu, K., Xu, S., Pearce, L., Wang, B., Unziff, M. L., Doganov, S., Chemaitilly, N., Superti-Furga, A., dos Santos, H., Migione, A., Morin, G., Gilleo, K., Heidmann, M., Schmitt, H., and Huber, C. (2011). Identification of an endogenous retroviral envelope gene with fucos- genic activity and placenta-specific expression in the rat: a new ‘paxton’ in a third order of mammals. Retrovirology 8, 67.

Huber, C., Diao-Suattageta, D., Glaess, A. O., Sullivan, J., Brauer, R., Wu, K., Xu, S., Pearce, L., Wang, B., Unziff, M. L., Doganov, S., Chemaitilly, N., Superti-Furga, A., dos Santos, H., Migione, A., Morin, G., Gilleo, K., Heidmann, M., Schmitt, H., and Huber, C. (2011). Identification of an endogenous retroviral envelope gene with fucos- genic activity and placenta-specific expression in the rat: a new ‘paxton’ in a third order of mammals. Retrovirology 8, 67.

Huber, C., Diao-Suattageta, D., Glaess, A. O., Sullivan, J., Brauer, R., Wu, K., Xu, S., Pearce, L., Wang, B., Unziff, M. L., Doganov, S., Chemaitilly, N., Superti-Furga, A., dos Santos, H., Migione, A., Morin, G., Gilleo, K., Heidmann, M., Schmitt, H., and Huber, C. (2011). Identification of an endogenous retroviral envelope gene with fucos- genic activity and placenta-specific expression in the rat: a new ‘paxton’ in a third order of mammals. Retrovirology 8, 67.

Huang, X., Yu, Y., Yang, H. W., Agar, J., Jung, Y. T., Lyu, M. S., Buckler-White, A., Huber, C., Dias-Santagata, D., Glaser, K., Hiom, K., Mele, M., and Gellert, M. (2011). Retrotransposon silencing by DNA methylation contributes to the evolution of placation and genome imprinting in mammals. Dev. Growth Differ. 53, 333–345.

Kaneko-Ishino, T., Kojima, T., Oono, R., and Ishino, F. (2006). Characterization of a molecularly cloned retrovirus in glioma cell lines. Cancer Gene Mol. 15, 37–49.

Kawasaki, H., Ebihara, T., Hata, K., Chiba, M., Katahira-Tayama, M., Morita, K., Miyachi, Y., and Kaneko-Ishino, T. (2006). Mouse homo sapiens of gene located on human chromosome 7q21. Proc. Natl. Acad. Sci. U. S. A. 103, 11258–11263.

Keller, K. J., Munnich, A., Pan, Z. Q., Win- ter, K., and Cormier-Daire, V. (2005). Characterization of a polytropic murine retrovirus. J. Virol. 79, 783–803.

Kim, S., Hwang, J., Yoon, H., Kim, S., Yoon, J., and Kim, D. (2010). A retroposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. Genome 53, 230–235.

Kimura, M. (1983). The Neutral The- ory of Molecular Evolution. Cambridge University Press, Cambridge, UK.

Kong, A. and Kim, S. (2012). A summary of the human imprinted gene landscape of skin-specific retroviral-like elements. Retrovirology 9, 37.

Koyama, H., Kobayashi, S., Aisaka, K., Nakamura, Y., Fujiyama, A., Shaw, G., Renfree, M. B., Kaneko-Ishino, T., and Ishino, F. (2001). A retroposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. Genome 53, 230–235.

Kosaki, K., Ferguson-Smith, A. C., and Surani, M. A. (1992). Targeted mutation of the Drosophila家 telomerase gene results in embryonic lethality. EMBO J. 11, 3337–3344.

Kotani, K., Kotsuka, K., Nakamura, Y., and Nakamura, Y. (2004). Coordinate downregulation of a novel imprinted transcript, TJP2, with PEG10 to placenta cells. J. Biol. Chem. 281, 17692–17696.

Kwan, C. L., Dagoneau, N., Chemaitilly, W., Xu, X., Pearce, K., Wang, R., Uzielli, G., Angel, P., and Hess, E. (2011). Mutator transposase is widespread in embryonic stem cells requires the histone deacetylase activity and placenta-specific expression. Placenta 31(Suppl.), 24, S21–S26.

Kawasaki, H., Ebihara, T., Hata, K., Chiba, M., Katahira-Tayama, M., Morita, K., Miyachi, Y., and Kaneko-Ishino, T. (2006). Mouse homolog of skin-specific retroviral-like aquatic prototaxins involved in skin development. J. Biol. Chem. 281, 27511–27525.

Kawasaki, H., Ebihara, T., Hata, K., Chiba, M., Katahira-Tayama, M., Morita, K., Miyachi, Y., and Kaneko-Ishino, T. (2006). Mouse homolog of skin-specific retroviral-like aquatic prototaxins involved in skin development. J. Biol. Chem. 281, 27511–27525.

Kawasaki, H., Ebihara, T., Hata, K., Chiba, M., Katahira-Tayama, M., Morita, K., Miyachi, Y., and Kaneko-Ishino, T. (2006). Mouse homolog of skin-specific retroviral-like aquatic prototaxins involved in skin development. J. Biol. Chem. 281, 27511–27525.

Kawasaki, H., Ebihara, T., Hata, K., Chiba, M., Katahira-Tayama, M., Morita, K., Miyachi, Y., and Kaneko-Ishino, T. (2006). Mouse homolog of skin-specific retroviral-like aquatic prototaxins involved in skin development. J. Biol. Chem. 281, 27511–27525.

Kawasaki, H., Ebihara, T., Hata, K., Chiba, M., Katahira-Tayama, M., Morita, K., Miyachi, Y., and Kaneko-Ishino, T. (2006). Mouse homolog of skin-specific retroviral-like aquatic prototaxins involved in skin development. J. Biol. Chem. 281, 27511–27525.

Kawasaki, H., Ebihara, T., Hata, K., Chiba, M., Katahira-Tayama, M., Morita, K., Miyachi, Y., and Kaneko-Ishino, T. (2006). Mouse homolog of skin-specific retroviral-like aquatic prototaxins involved in skin development. J. Biol. Chem. 281, 27511–27525.

Kawasaki, H., Ebihara, T., Hata, K., Chiba, M., Katahira-Tayama, M., Morita, K., Miyachi, Y., and Kaneko-Ishino, T. (2006). Mouse homolog of skin-specific retroviral-like aquatic prototaxins involved in skin development. J. Biol. Chem. 281, 27511–27525.
Kaneko-Ishino and Ishino Domesticated genes in mammals

Schmidt, D., Schwalie, P. C., Wicks, F., and Ohlm, D. T. (2012). Waves of retrotransposon expansion remedied genome organization and CTCF binding in multiple mammalian lineages. Cell 148, 355–368.

Smit, A. F. (1999). Interspersed repeats and characterisation of a developing imprinted gene, Rtl1 site-targeted silencing of murine leukemia virus. J. Virol. 73, 126–134.

Stoye, J. P. (2006). Rtl1 endogenous koala retrovirus. Nature 441, 1201–1204.

Tarlinton, R., Meers, J., and Young, P. (2004). Role of retrotransposon-derived imprinted genes, R75, in the fetal-maternal interface of mouse placententa. Nat. Genet. 40, 243–248.

Ungemach, K., Bremman, J., Wall, E., Watson, C. J., Stott, D., Rigby, P. W., and Reith, A. D. (2001). Identification and characterization of a developmentally regulated mammalian gene that utilizes a programmed somatic frameshifting. Nucleic Acids Res. 29, 4797–4806.

Watanabe, S., Noel, E., and Sasaki, H. (2006). Evolution of the vertebrate IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. 38, 1788–1790.

Wolf, D., and Golub, S. P. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. Cell 131, 46–57.

Wolf, D., and Golub, S. P. (2004). Embryonic stem cells use ZFP809 to silence retroviral DNA. Nature 458, 1201–1204.

Yokomine, T., Hata, K., Tsuda, M., and Sasaki, H. (2006). Evolution of the vertebrate IAP retrovirus gene family: a possible link between existence of IAP-like and genomic imprinting. Cytogenet. Genome Res. 115, 75–80.

Copyright © 2012 Kaneko-Ishino and Ishino. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution, and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics, etc.