DNA Research **18**, 211–219, (2011)
Advance Access Publication on 2 June 2011
doi:10.1093/dnares/dsr012

# Identification of tammar wallaby *SIRH12*, derived from a marsupial-specific retrotransposition event

Ryuichi Ono[1], Yoko Kuroki[2], Mie Naruse[1,3], Masayuki Ishii[1,2], Sawa Iwasaki[1], Atsushi Toyoda[4], Asao Fujiyama[4,5], Geoff Shaw[6,7], Marilyn B. Renfree[6,7], Tomoko Kaneko-Ishino[3], and Fumitoshi Ishino[1,*]

*Department of Epigenetics, Medical Research Institute, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan[1]; Laboratory for Immunogenomics, RIKEN Research Center for Allergy and Immunology, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa 230-0045, Japan[2]; School of Health Sciences, Tokai University, Bohseidai, Isehara, Kanagawa 259-1193, Japan[3]; Comparative Genomics Laboratory, National Institute of Genetics, Yata 1111, Mishima, Shizuoka 411-8540, Japan[4]; Department of Informatics, The Graduate University for Advanced Studies, and National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan[5]; Australian Research Council Centre of Excellence for Kangaroo Genomics, Victoria 3010, Australia[6] and Department of Zoology, The University of Melbourne, Victoria 3010, Australia[7]*

*To whom correspondence should be addressed. Tel./Fax. +81-3-5803-4863. Email: fishino.epgn@mri.tmd.ac.jp

Edited by Toshihiko Shiroishi
(Received 28 March 2011; accepted 1 May 2011)

**Abstract**

**In humans and mice, there are 11 genes derived from sushi-ichi related retrotransposons, some of which are known to play essential roles in placental development. Interestingly, this family of retrotransposons was thought to exist only in eutherian mammals, indicating their significant contributions to the eutherian evolution, but at least one, *PEG10*, is conserved between marsupials and eutherians. Here we report a novel sushi-ichi retrotransposon-derived gene, *SIRH12*, in the tammar wallaby, an Australian marsupial species of the kangaroo family. *SIRH12* encodes a protein highly homologous to the sushi-ichi retrotransposon Gag protein in the tammar wallaby, while *SIRH12* in the South American short-tailed grey opossum is a pseudogene degenerated by accumulation of multiple nonsense mutations. This suggests that *SIRH12* retrotransposition occurred only in the marsupial lineage but acquired and retained some as yet unidentified novel function, at least in the lineage of the tammar wallaby.**

**Key words:** retrotransposon; evolution of mammals; marsupials

## 1. Introduction

About 40% of our genome is derived from transposable elements, such as DNA transposons and retrotransposons.[1–3] They have long been considered as junk DNAs. However, it has become clear that certain genes derived from retrotransposons play essential roles in mammalian development as newly acquired endogenous genes. We previously identified human and mouse *PEG10* (*Paternally expressed 10*) as a novel paternally expressed imprinted gene.[4,5] *PEG10* is a single-copy gene located between *SGCE* (*Sarcoglycan epsilon*) and *PPP1R9A* (*Protein phosphatase 1, regulatory* (*inhibitor*) *subunit 9A*). It is highly conserved in not only eutherian but also in marsupial mammals, but it is absent from monotreme mammals and from other vertebrates, such as birds and fish.[5,6] A structural analysis of *PEG10* clearly showed that it was derived from one of the sushi groups of Ty3/Gypsy LTR (long terminal repeat) retrotransposons. *PEG10* has two separate open

reading frames (ORF1 and ORF2) that produce proteins similar to Gag and Pol proteins of the sushi-ichi retrotransposon from fugu fish, and still retains a -1 frameshifting mechanism to produce Gag-Pol (ORF1 and ORF2) fusion protein, as is always seen in the Ty3/Gypsy retrotransposons and retroviruses.[4,7–9] *Peg10* deficient mice, which lack both ORF1 and ORF2, have early embryonic lethality before 10.5 days post-coitus owing to severe placental defects.[10] Similarly, *Peg11/Rtl1*, another retrotransposon-derived imprinted gene highly conserved in the eutherian species, also plays an essential role in the maintenance of the placenta during pregnancy.[11] Its loss leads to late fetal/neonatal lethality because of placental malfunction. *PEG11* retrotransposon insertion occurred before divergence of the eutherians and marsupials, but *PEG11* became degenerated in the marsupial lineage.[12] Therefore, *PEG11* is a eutherian-specific *SIRH* (*Sushi-Ichi Retrotransposon Homologues*) family gene, which is critical for the maintenance of the normal placental structure and function during the middle and late embryonic period in mice.[11,13]

There are 11 genes that are similar to the sushi-ichi retrotransposon (*SIRH* family genes, also called *MART* or *SUSHI* genes), including *PEG10* and *PEG11/RTL1* in humans and mice, and they are highly conserved in the eutherian mammals.[4,5,8,10,14–17] As mentioned above, *PEG10/SIRH1* and *PEG11/SIRH2* are paternally expressed imprinted genes on human chromosome 7q21/mouse proximal chromosome 6 and human chromosome 14q32/mouse distal chromosome 12, respectively.[13,18] In mice, *Sirh3/Ldoc1l* is another autosomal gene showing biallelic expression on the distal chromosome 15, while *Sirh4*, *Sirh5*, *Sirh6*, *Sirh7*, *Sirh8*, *Sirh9*, *Sirh10* and *Sirh11* are located on the X chromosome. To elucidate functions of all the *SIRH* family genes, systematic production of KO (knock out) mice for all the *SIRH* family genes are now in the process. It is possible that the other *SIRH* family genes have essential functions like *PEG10/SIRH1* and *PEG11/SIRH2*.

With the recent availability of marsupial genome data, it is possible to screen marsupial-specific *SIRH* family genes by comparing them with those of eutherians and those of other vertebrate genomes. Here we report the identification of *SIRH12* as a novel retrotransposon-derived gene in tammar wallaby.

## 2. Materials and Methods

### 2.1. Animals and tissue collection

Tammar wallabies of Kangaroo Island origin were maintained in the University of Melbourne marsupial breeding colony in grassy, outdoor enclosures. Lucerne cubes, grass and water were provided ad libitum and supplemented with fresh vegetables. Fetuses and yolk sac placentas were collected between days 23 and 25 of the 26.5-d gestation.[19,20] A series of tissues were collected from pouch young [d152 *post partum* (pp)]. Experimental procedures conformed to Australian National Health and Medical Research Council (1990) guidelines and were approved by the Animal Experimentation Ethics Committees of the University of Melbourne.

### 2.2. RT-PCR and 5′- and 3′-RACE

Genomic DNA and total RNA were prepared from fetuses and yolk sac placentas from wallaby conceptuses and several tissues from a pouch young, using TRISOL (Invitrogen), as described in manufacturer's protocol. cDNA was synthesized from 1 μg of total RNA using Superscript III reverse transcriptase (Invitrogen) with an oligo dT primer. Gene expression profiles were deduced by agarose gel electrophoresis of RT-PCR products with ethidium bromide (EtBr) staining. The primers used for the expression profiles were as follows: *SIRH12*-F (5′-TTTCTCCAGCTGTTCTGGCT-3′), *SIRH12*-R (5′-CAGGGTAGAGGGGAGGTTTC-3′), *GAPDH*-F (5′- AGAAAGTGGTGAAGCAGGCAT-3′) and *GAPDH*-R (5′-TGGAGGACATGTAGACCATGAG-3′). RACE reactions were performed with wallaby liver and large intestine using RNA SMARTER RACE cDNA Amplification kit (Clontech) according to the manufacturer's recommendations. The 5′- and 3′-RACE fragments were generated with the gene-specific primers *SIRH12*-5′-RACE (5′-TCCATGTGGCCAAGTTCTGAGGATTC-3′) and *SIRH12*-3′-RACE (5′-GAATCCTCAGAACTTGGCCACATGGA-3′), respectively. PCR conditions were as described previously.[5]

### 2.3. Comparative genomics analysis

Identification of *SIRH* family genes was performed using the TBLASTN and BLSTP program from NCBI server (http://www.ncbi.nlm.nih.gov/BLAST/) against eutherian and marsupial genomes using sushi-ichi Gag protein as a query (GenBank ID. AF030881). Sequence analysis was performed using the ClustalW program (http://clustalw.ddbj.nig.ac.jp/top-j.html). The sequences of opossum *SIRH12* syntenic region [6993247–8893247 *Monodelphis domestica* chromosome 3 genomic contig, reference assembly (based on MonDom5)], tammar wallaby *Macropus eugenii SIRH12* region, mouse *Sirh12* syntenic region (5099338–6288585 *Mus musculus* strain C57BL/6J chromosome 13 genomic contig, MGSCv37), human *SIRH12* syntenic region (c24594359–23494359 *Homo sapiens* chromosome 5 genomic contig, GRCh37.p2 reference primary assembly), platypus (12010399–12780399 *Ornithorhynchus anatinus* chromosome 1 genomic contig, reference assembly [based on

**Table 1.** List of SIRH family gene candidates

| Gene name | SIRH number | Other alias | Accession number | Expect | Identity |
|---|---|---|---|---|---|
| A | | | | | |
| PEG10 | SIRH1 | EDR, HB-1, KIAA1051, MEF3L, Mar2, Mart2, RGAG3 | NM_001040152 | 3.00E−24 | 97/359 (28) |
| RTL1 | SIRH2 | MART1, Mar1, PEG11 | NM_106713 | 3.00E−18 | 63/198 (32) |
| LDOC1L | SIRH3 | DKFZp761O17121, Mar6, Mart6, dJ1033E15.2 | NM_032287 | 7.00E−04 | 28/88 (32) |
| FAM127C | SIRH4 | RP4-809E13.1, CXX1c, FLJ25577, MAR8B | NM_001078173 | 0.45 | 26/76 (35) |
| FAM127A | SIRH5 | CXX1, MAR8C, MART8C, MGC117411, Mar8, Mart8 | NM_001078171 | 0.35 | 30/93 (33) |
| FAM127B | SIRH6 | CXX1b, DKFZp564B147, MAR8A, MGC8471 | NM_001078172 | 0.34 | 33/111 (30) |
| LDOC1 | SIRH7 | BCUR1, Mar7, Mart7 | NM_012317 | 0.002 | 24/76 (32) |
| RGAG4 | SIRH8 | RP11-262D11.3, 6430402L03Rik, KIAA2001, MAR5, MART5 | NM_001024455 | 1.00E−09 | 41/157 (27) |
| ZCCHC5 | SIRH9 | FLJ38865, Mar3, Mart3, ZHC5 | NM_152694 | 3.00E−12 | 63/254 (25) |
| RGAG1 | SIRH10 | KIAA1318, MAR9, MART9, MGC142230 | NM_020769 | 6.00E−07 | 39/137 (29) |
| ZCCHC16 | SIRH11 | FLJ46608, Mar4, Mart4 | NM_001004308 | 1.00E−05 | 57/242 (24) |
| B | | | | | |
| Peg10 | Sirh1 | AA407948, Edr, HB-1, MEF3L, Mar2, Mart2, MyEF-3 | NM_001040611 | 2.00E−28 | 81/253 (33) |
| Rtl1 | Sirh2 | 6430411K18Rik, Mar, Mart1, Mor1, Peg11 | NM_184109 | 4.00E−17 | 56/166 (34) |
| Ldoc1l | Sirh3 | BC058638, MGC73499, Mar6, Mart6, sushi-15E3 | NM_177630 | 5.00E−04 | 51/202 (26) |
| CAAX box 1 homolog C | Sirh4 | RP23-479D16.1, 2900027G03Rik, Mar8.1, Mart8a | NM_028375 | 0.003 | 26/76 (35) |
| CAAX box 1 homolog A | Sirh5 | Mart8b; Mar8.2A/B; 1110012O05Rik; | NM_024170 | 0.003 | 26/76 (35) |
| CAAX box 1 homolog B | Sirh6 | Mart8c | NM_001018063 | 0.003 | 26/76 (35) |
| Ldoc1 | Sirh7 | RP23-322K17.2, Gm366, Mar7, Mart7 | NM_001018087 | 0.81 | 22/76 (29) |
| Rgag4 | Sirh8 | RP23-448C18.4, 6430402L03Rik, KIAA2001, Mar5, Mart5, mKIAA2001, sushi-XC3 | NM_183318 | 2.00E−10 | 4/167 (27) |
| Zcchc5 | Sirh9 | RP23-233G6.4, D430021I08Rik, Gm375, Mar3, Mart3, ZHC5, sushi-XD | NM_199468 | 2.00E−09 | 63/258 (25) |
| Rgag1 | Sirh10 | RP23-71M18.1, Gm385, KIAA1318, Mar9, Mart9, mKIAA1318, sushi-XF2 | NM_001040434 | 3.00E−09 | 42/137 (31) |
| Zcchc16 | Sirh11 | RP23-319K12.1, C230031A03Rik, Mar4, sushi-XF2b | NM_001033795 | 2.00E−07 | 60/283 (22) |
| C | | | | | |
| PEG10 | SIRH1 | | ABQO010716413 | 2.00E−18 | 88/317 (27) |
| Degenerated | | | ABQO010379794 | 0.008 | 30/95 (31) |
| Degenerated | | | ABQO010296533 | 0.03 | 40/140 (28) |
| Degenerated | | | ABQO010214722 | 0.39 | 18/60 (30) |
| SIRH12 | SIRH12 | | ABQO010016898 | 7.00E−05 | 16/52 (31) |

A, Human genes; B, mouse genes and C, wallaby genes. Values in parenthesis are percentages values.

*Note*. Analysis was performed using the TBLASTN and BLAST program from NCBI server (http://www.ncbi.nlm.nih.gov/BLAST/) against human, mice, opossum and wallaby genomes using sushi-ichi Gag protein as a query (GenBank ID: AF030881).

*Ornithorhynchus_anatinus*-5.0.1)], chicken (930001–1330000 *Gallus gallus* chromosome Z genomic contig, reference assembly [based on *Gallus_gallus*-2.1)], Frog (c1800000–1200001 *Xenopus* (Silurana) *tropicalis* unplaced genomic scaffold, v4.2 XENTRscaffold_113) and *Fugu* (FUGU4:scaffold_49:800001:930000:-1) were extracted from NCBI (http://www.ncbi.nlm.nih.gov/). The tammar BAC (Bacterial artificial chromosome) library (MEB1) were screened using tammar *SIRH12* sequence as a probe by PCR at the RIKEN Yokohama Institute. Sequence of BAC clone (MEB1-141D12), which includes *SIRH12* was determined at the National Institute of Genetics by a combined shotgun/nested deletion strategy adopted to sequence

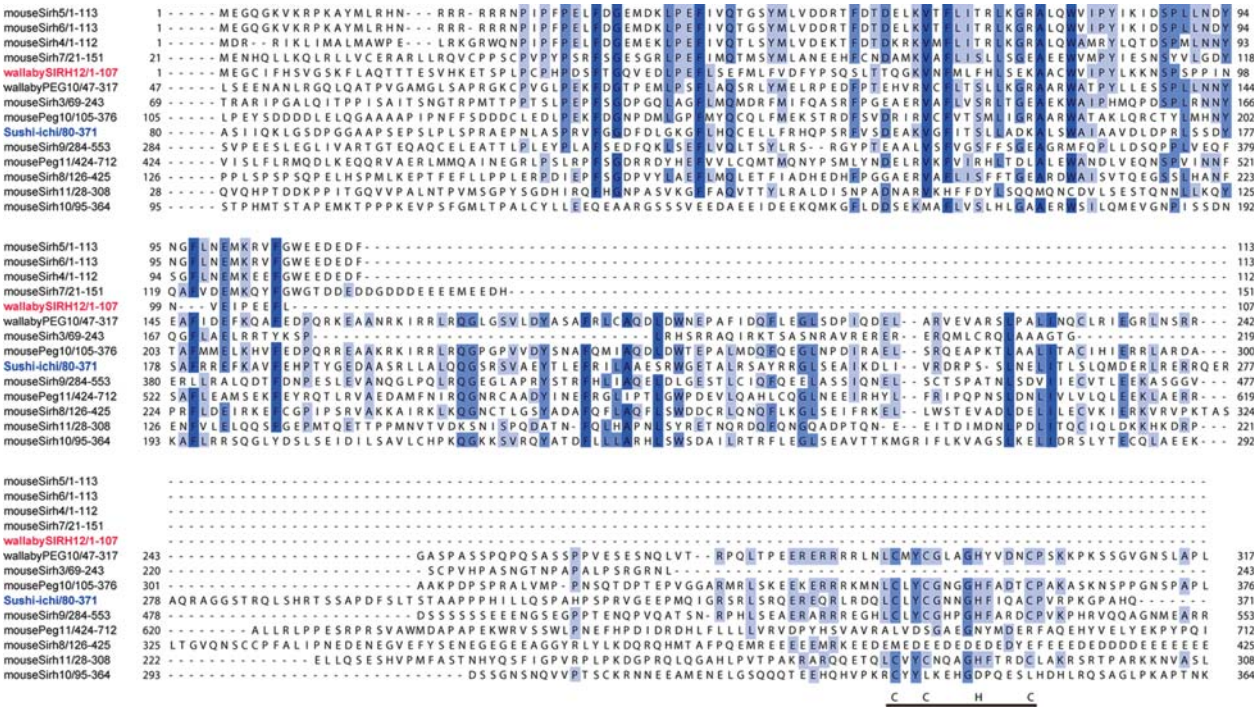


**Figure 1.** Amino acid alignment of SIRH family genes. Alignment of the amino acid sequence of the Gag-like regions of tammar *SIRH12* and *PEG10*, mouse *Sirh* family genes and Gag region of sushi-ichi retrotransposon from Fugu fish is presented. CX2CX4HX4C zinc finger motif conserved in Ty3/Gypsy type retrotransposons is indicated. Highly conserved residues are in dark blue and relatively identical residues are in light blue.

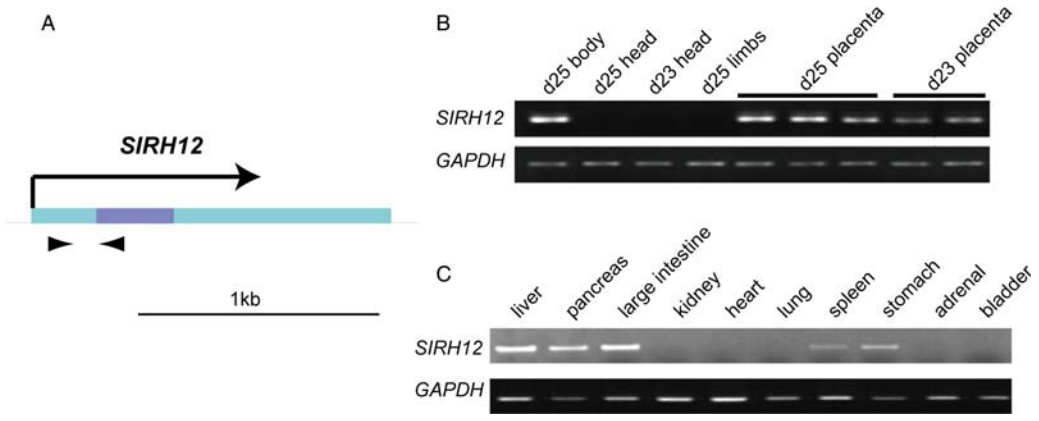


**Figure 2.** Genomic structure and expression profiles of *SIRH12*. (A) Genomic structure of full-length tammar *SIRH12*. An arrow represents the direction of *SIRH12* gene. UTR (untranslated region) and ORF (open reading frame) are indicated by blue and purple boxes, respectively. The primer positions used for RT-PCR are indicated by arrowheads. (B) Expression profiles of *SIRH12* in the tammar fetus and yolk sac placenta (between days 23 and 25 of pregnancy). The RT-PCR products using total RNA from the tammar fetus and yolk sac placenta are shown. Expression of tammar *GAPDH* (*glyceraldehyde-3-phosphate dehydrogenase*) for each sample is shown as a control. (C) Expression profiles of *SIRH12* in several tissues of wallaby pouch young (d152 pp). The RT-PCR products using total RNA from several tissues of wallaby pouch young are shown. Expression of tammar *GAPDH* for each sample is shown as a control.

the BAC inserts as described previously.[21–24] The primer sequences and conditions for their use are available on request. The sequence data have been submitted to GenBank (http://www.ncbi.nlm.nih.gov/genbank/) under GenBank ID. JF720345. RepeatMasker (http://www.repeatmasker.org) was used for the detection of long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs) and LTR elements in the genomic region. Alignments were obtained using the VISTA Web server (http://www-gsd.lbl.gov/VISTA). *SIRH12* syntenic regions of several species identified above were aligned using the default setting (>70% identity and >100 bp in length) of mVISTA with the LAGAN program.

## 3. Results

### 3.1. Identification of a SIRH family gene in marsupials

Sushi-ichi is the first full-length vertebrate LTR retrotransposon isolated from Fugu fish. The sushi-ichi Gag protein has a unique amino acid sequence among the LTR retrotransposons, so sushi-ichi retrotransposon homologues (*SIRH*) were screened by TBLASTN and BLASTP analyses. In humans and mice, 11 *SIRH* family

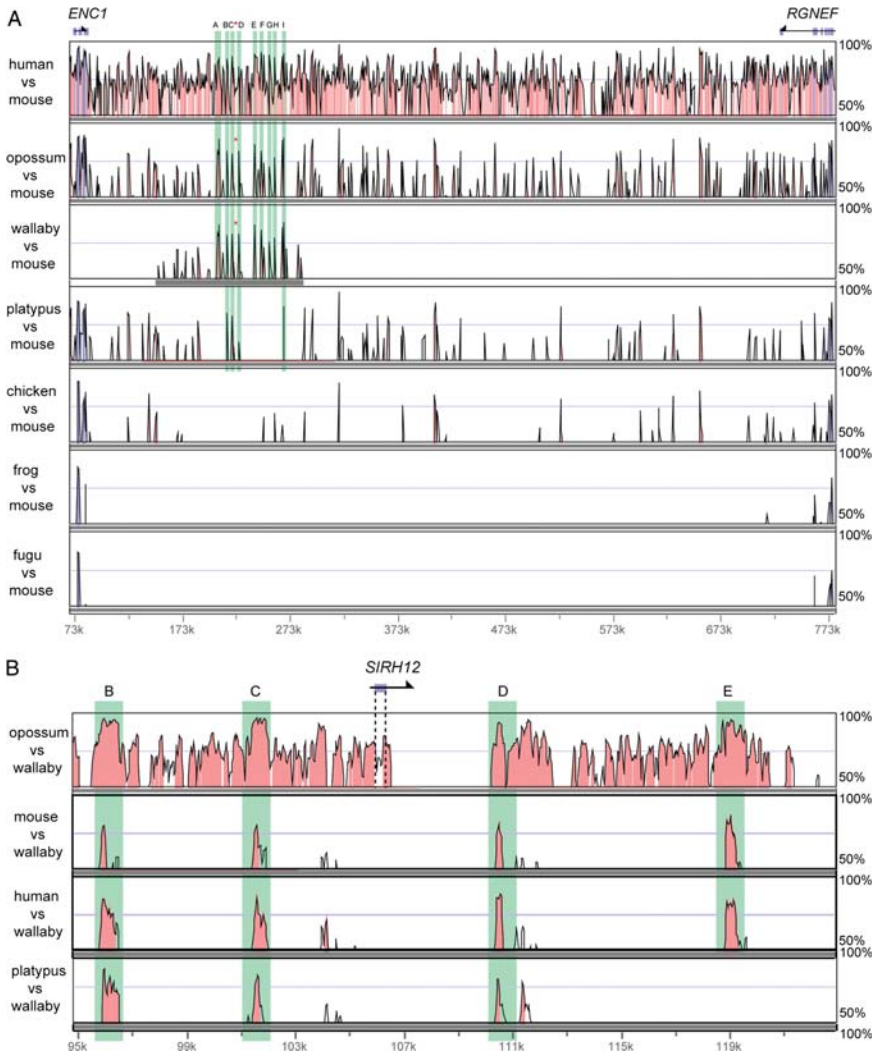


**Figure 3.** Mapping and comparative genomic analysis of the tammar wallaby *SIRH12*. (A) Mapping of the tammar wallaby BAC including *SIRH12* between *ENC1* and *RGNEF* gene. mLAGAN alignment of mouse, human, opossum, platypus, chicken, frog and fugu *ENC1*/*RGNEF* region and wallaby BAC sequence including *SIRH12* are shown based on the mouse sequence using mVISTA program. Default parameters for mVISTA were used (conserved level, 70%). Conserved regions appear as peaks highlighted in pink (~70% identity). When these regions coincide with exonic sequences of ENC1 or RGNEF, the peaks are shaded in purple. There are nine ECRs (evolutionary conserved regions) in the tammar BAC sequence (ECRs A–I; green boxes) between mice and opossum in a region orthologous to *ENC1*/*RGNEF*. Tammar wallaby and opossum *SIRH12* are located between ECRs C and D, represented by red stars. (B) Conserved location of *SIRH12* in marsupials but not in eutherians. mLAGAN alignment of opossum, mouse, human and platypus between ECRs B and E based on wallaby BAC sequences. The black arrow represents wallaby *SIRH12* located between ECRs C and D. *SIRH12* orthologous sequence in opossum was conserved but its protein-coding frame had degenerated by accumulation of nonsense mutations. There is no evidence of *SIRH12* in eutherian mammals, monotremes, chicken and fugu.

genes have been identified previously.[4,5,8,10,14–17] (Table 1A, B). In two marsupials (the grey short-tailed opossum and tammar wallaby), five sequences were identified in tammar wallaby (Table 1C) and none in the grey short-tailed opossum. One corresponded to tammar *PEG10* as previously reported.[6] Another sequence (GenBank ID. ABQO010016898) also shared high homology in the amino acid sequence with other *SIRH* family genes, therefore, we named it *SIRH12*, as a novel candidate in the *SIRH* gene family (Fig. 1). The remaining three sequences seemed to be non-functional because their open reading frames had accumulated multiple nonsense mutations.

### 3.2. *Genomic structure and expression of SIRH12 in tammar wallaby pouch young tissues*

*SIRH12* full-length sequence consisting of 1492 bp was determined by 5′ RACE (Rapid Amplification of cDNA Ends) and 3′-RACE (GenBank ID. JF710845). *SIRH12* is an intron-less gene, as are most *SIRH* family genes, such as *SIRH3*, *SIRH4*, *SIRH5*, *SIRH6*, *SIRH7*, *SIRH8* and *SIRH10*. *SIRH12* has a candidate ORF consisting of 107 amino acids showing high similarity with the Gag protein of sushi-ichi (Fig. 2A). However, it lacks CCHC zinc finger motif, conserved in the Gag protein in retrotransposons and retroviruses, and a part corresponding with Pol protein that is retained in *PEG10*, *PEG11/RTL1* and *SIRH9*. LTR sequences that are usually attached at both ends of the retrotransposons and retroviruses are not recognizable as there are no terminal redundancy sequences and no homology with existing LTR sequences within both 5 kb sequences of upstream and downstream of *SIRH12*. Therefore, *SIRH12* may be transcribed from a promoter that was derived from host genome. However, it is also possible that a promoter on the *SIRH12* LTR has degenerated during evolution but still drives *SIRH12* transcription. All these data suggest that *SIRH12* has already lost its retotranspositional ability.

Expression of *SIRH12* was analysed in wallaby fetuses and yolk sac placentas (Fig. 2B), and several tissues from a day 152 wallaby pouch young (Fig. 2C). *SIRH12* was not expressed in head and limb but was expressed in the bodies of the tammar fetuses and yolk sac placentas while, in pouch young tissues, *SIRH12* was expressed in the endoderm-derived tissues, liver, pancreas, large intestine, spleen and stomach, but was not detected in the lung nor in the tissues primarily of mesodermal and ectodermal origin, namely kidney, heart, adrenal and bladder. This endodermal expression pattern in pouch young is unique compared with those of other *SIRH* family genes in the eutherian mammals, while expression in the placenta is relatively common in *SIRH* family genes.

### 3.3. *Comparative analysis of SIRH12*

To elucidate whether *SIRH12* is an orthologue for either one of the *SIRH1-11* genes in humans and mice, precise mapping of tammar wallaby *SIRH12*

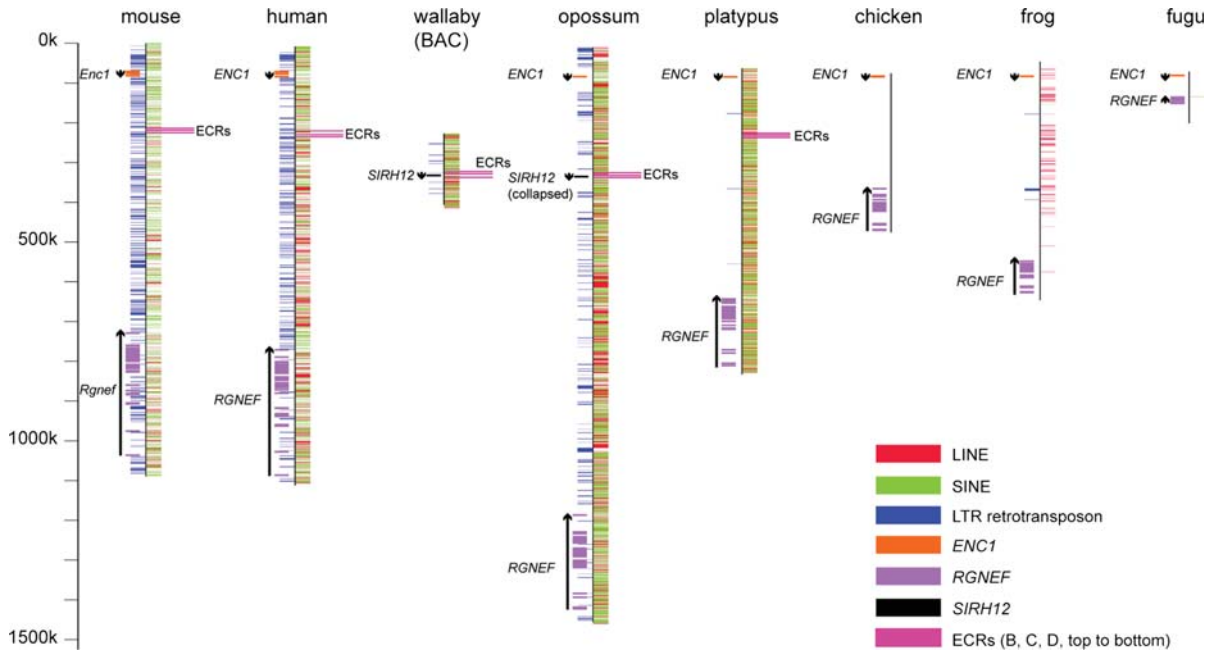


**Figure 4.** Retrotransposons between *ENC1* and *RGNEF* regions and chromosomal location of wallaby *SIRH12*. Red and green bars on the right side of the center lines represent LINE and SINE sequences, respectively, and blue bars on the left side represent LTR retrotransposons. Orange, purple and black bars on the left side represent *ENC1*, *RGNEF* and *SIRH12*, respectively. Long pink bars on the right side represent ECRs B, C and D from top to bottom. *SIRH12* is located between ECRs C and D.

was carried out by sequencing a wallaby BAC clone containing *SIRH12*. *SIRH12* was located between two neighbouring genes, *ENC1* (ectodermal-neural cortex) and *RGNEF* (Rho-guanine nucleotide exchange factor) that are conserved in vertebrates. In the BAC sequence, there are several syntenic regions, so-called evolutionary conserved regions (ECRs A–I in Fig. 3A), in the interval between *ENC1* and *RGNEF*. ECRs B, C, D and I are conserved among all three mammalian groups including the egg-laying mammals (the monotreme platypus), while ECRs A, E, F, G and H are conserved only in therian mammals (the eutherians and marsupials). It is clear that tammar *SIRH12* is located between ECRs C and D and that a *SIRH12* orthologue in the South American opossum resides in the same location, although it is degenerated and does not have a long ORF corresponding to the wallaby *SIRH12* (Fig. 3B). Importantly, in eutherian mammals there are no *SIRH12* orthologues present between the ECRs C and D. The same is true of the platypus, chicken, frog and fish, indicating that *SIRH12* retrotransposition occurred only in the marsupial lineage after their divergence from the eutherian mammals that occurred between 130 and 148 million years ago.[25,26]

### 3.4. *Evolutions of SIRH family genes*

Using published sequences, we compared the entire region between *ENC1* and *RGNEF* among several vertebrates from fish (fugu) to mammals. As is reported, the size of this region is the smallest in fugu fish and that of opossum is the largest. There are numerous LINEs and SINEs (red and green bars in Fig. 4) in all mammalian groups. By insertions of these elements mammalian genomes become longer than those of fugu fish and chicken. Consistent with the previous report, LTR-type retrotransposons are absent in the platypus[27] (blue bars in Fig. 4). The *PEG10* retrotransposon insertion occurred in the genome of the therian ancestor and was incorporated into the genomes of both marsupials and eutherians[6] (Fig. 5A), while *PEG11/RTL1* is a eutherian-specific gene[12] (Fig. 5B). It is highly likely that other eutherian *SIRH* family genes, *SIRH3* to *SIRH11*, are not present in the marsupials (M. Naruse, M. Ishii and R. Ono, unpublished data), suggesting that their retrotranspositions occurred only in the eutherian lineage. Our data in this report indicate that the original *SIRH12* was retrotransposed into the ancestral marsupial genome after the eutherian–marsupial divergence and became incorporated into the tammar wallaby genome but degenerated in that of the opossum (Fig. 5C).

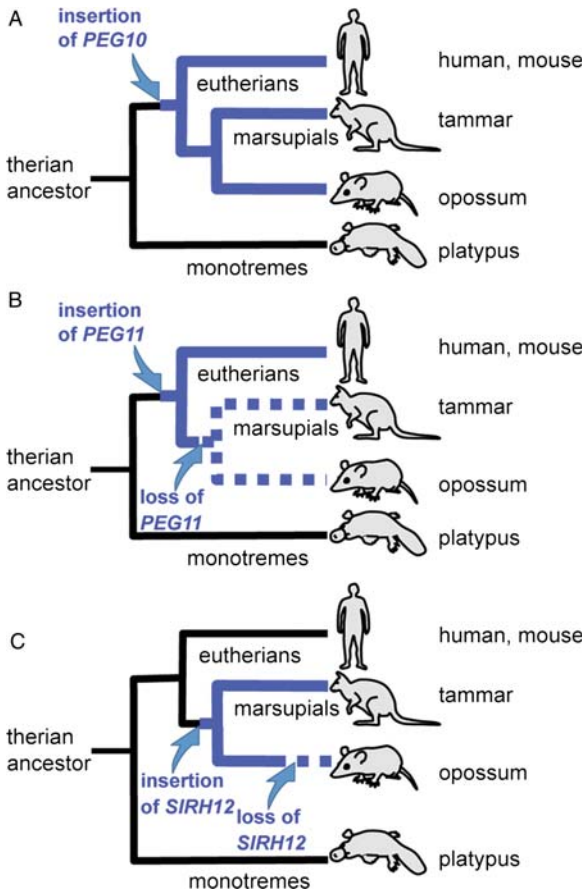


**Figure 5.** Possible evolutionary pathway of the *SIRH* family genes in mammals. (A) *PEG10* insertion occurred in a therian ancestor and domesticated before the split of marsupials and eutherians. (B) *PEG11/RTL1* insertion occurred in a therian ancestor but domesticated only in the eutherians and collapsed in the marsupials. (C) *SIRH12* insertion occurred in a marsupial ancestor and domesticated at least in wallaby but collapsed in opossum.

## 4. Discussion

In this study, we identified a novel sushi-ichi retrotransposon-derived gene, *SIRH12*, in the tammar wallaby, an Australian marsupial species of the kangaroo family. Comparative genomic analysis suggests that *SIRH12* is present in the marsupial lineage but is not present in the eutherian and monotreme lineages. Together with other *SIRH* genes, *PEG10*, *PEG11/RTL1* and *SIRH3-11*, it is probable that the sushi-ichi-like retrotransposons were once active and retrotransposed around the time of the divergence between marsupials and eutherians, contributing to the evolution of both the extant eutherian and marsupial mammals. In general, as retrotransposons are harmful for host animals, they tend to be inactivated by DNA methylation, nonsense and frame-shift mutations.[28] However, some were incorporated into their genomes and became functional, so were selected positively, presumably because they were

advantageous to their host animals.[10,11,29–33] Although it remains unknown when the opossum *SIRH12* degenerated after its incorporation in the marsupial lineage, *SIRH12* may be functional at least in the wallaby because its protein-coding frame has been maintained and is actively transcribed in several tissues. As species-specific genes are strong candidates for species-specific functions, it would be interesting to know whether *SIRH12* has a specific function. The opossum and the tammar are very different marsupials. The tammar is a macropodid marsupial of the highly evolved kangaroo family, but the grey short-tailed opossum is a generalized marsupial closest in form to the ancestral marsupials. The tammar placenta is more highly specialized than that of the opossum[34] and there are many structural and functional differences between tammar and opossum placentas.[35–37] Although both *Peg10* and *Peg11* are derived from the sushi-ichi retrotransposon and have Gag and Pol-like sequences, they have biologically very different functions: the former has a function in the formation of spongiotrophoblast and labyrinth layer, while the latter has a function in the maintenance of feto-maternal interface.[10,11] Therefore, *SIRH* family genes could have a wide variety of functions not only in the placenta but in some other organs to perform eutherian- or marsupial-specific functions. Because tammar *SIRH12* is expressed in the yolk sac placenta, it is therefore possible that *SIRH12* fulfils some role in tammar placentation.

**Acknowledgements:** We are grateful to all the technical staff of the Sequence Technology Team at RIKEN Genomic Sciences Center (GSC) and Comparative Genomics Laboratory (National Institute of Genetics) for their assistance. We thank Scott Brownlees, Helen Clark and the wallaby research group for assistance with the animals and tissue collections.

## Funding

This work was supported in part by Grants-in-Aid for Scientific Research on Priority Area (R.O.) and Innovative Areas 'Genome Science' (R.O., Y.K., A.T.) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, Grant-in-Aid for Young Scientist (R.O.), Creative Science Research (F.I. and T.K.-I.), and Bilateral Program on Joint Research Project (F.I.) from the Japan Society for the Promotion of Science (J.S.P.S.), grants from Mochida Memorial Foundation for Medical and Pharmaceutical Research and The Nakajima Foundation (R.O.). M.B.R. was supported by an Australian Research Council Federation Fellowship, and G.S. and M.B.R. by the ARC Centre of Excellence in Kangaroo Genomics.

## References

1. Lander, E., Linton, L., Birren, B., et al. 2001, Initial sequencing and analysis of human genome, *Nature*, **409**, 860–921.
2. Waterston, R.H., Lindblad-Toh, K., Birney, E., et al. 2002, Initial sequencing and comparative analysis of the mouse genome, *Nature*, **420**, 520–62.
3. Gibbs, R.A., Weinstock, G.M., Metzker, M.L., et al. 2004, Genome sequence of the Brown Norway rat yields insights into mammalian evolution, *Nature*, **428**, 493–521.
4. Ono, R., Kobayashi, S., Wagatsuma, H., et al. 2001, A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21, *Genomics*, **73**, 232–7.
5. Ono, R., Shiura, H., Aburatani, H., Kohda, T., Kaneko-Ishino, T. and Ishino, F. 2003, Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6, *Genome Res.*, **13**, 1696–705.
6. Suzuki, S., Ono, R., Narita, T., et al. 2007, Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting, *PLoS Genet.*, **3**, e55.
7. Poulter, R. and Butler, M. 1998, A retrotransposon family from the pufferfish (fugu) Fugu rubripes, *Gene*, **215**, 241–9.
8. Volff, J., Korting, C. and Schartl, M. 2001, Ty3/Gypsy retrotransposon fossils in mammalian genomes: did they evolve into new cellular functions?, *Mol. Biol. Evol.*, **18**, 266–70.
9. Shigemoto, K., Brennan, J., Walls, E., et al. 2001, Identification and characterisation of a developmentally regulated mammalian gene that utilises -1 programmed ribosomal frameshifting, *Nucleic Acids Res.*, **29**, 4079–88.
10. Ono, R., Nakamura, K., Inoue, K., et al. 2006, Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality, *Nat. Genet.*, **38**, 101–6.
11. Sekita, Y., Wagatsuma, H., Nakamura, K., et al. 2008, Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta, *Nat. Genet.*, **40**, 243–8.
12. Edwards, C.A., Mungall, A.J., Matthews, L., et al. 2008, The evolution of the DLK1-DIO3 imprinted domain in mammals, *PLoS Biol.*, **6**, e135.
13. Kagami, M., Sekita, Y., Nishimura, G., et al. 2008, Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes, *Nat. Genet.*, **40**, 237–42.
14. Lynch, C. and Tristem, M. 2003, A co-opted gypsy-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats, *Curr. Biol.*, **13**, 1518–23.

15. Brandt, J., Veith, A.M. and Volff, J.N. 2005, A family of neofunctionalized Ty3/gypsy retrotransposon genes in mammalian genomes, *Cytogenet. Genome Res.*, **110**, 307–17.
16. Youngson, N.A., Kocialkowski, S., Peel, N. and 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**, 481–90.
17. Butler, M., Goodwin, T., Simpson, M., Singh, M. and Poulter, R. 2001, Vertebrate LTR retrotransposons of the Tf1/sushi group, *J. Mol. Evol.*, **52**, 260–74.
18. Seitz, H., Youngson, N., Lin, S.P., et al. 2003, Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene, *Nat. Genet.*, **34**, 261–2.
19. Renfree, M.B. and Tyndale-Biscoe, C.H. 1973, Intrauterine development after diapause in the marsupial Macropus eugenii, *Dev. Biol.*, **32**, 28–40.
20. Renfree, M.B. and Tyndale-Biscoe, C.H. 1978, Manipulation of marsupial embryos and pouch young. In: Daniel, J.C. (ed.). *Methods of Mammalian Reproduction*, Academic Press, New York, pp. 307–31.
21. Fleischmann, R.D., Adams, M.D., White, O., et al. 1995, Whole-genome random sequencing and assembly of Haemophilus influenzae Rd, *Science*, **269**, 496–512.
22. Hattori, M., Tsukahara, F., Furuhata, Y., et al. 1997, A novel method for making nested deletions and its application for sequencing of a 300 kb region of human APP locus, *Nucleic Acids Res.*, **25**, 1802–8.
23. Hattori, M., Fujiyama, A., Taylor, T.D., et al. 2000, The DNA sequence of human chromosome 21, *Nature*, **405**, 311–9.
24. Kuroki, Y., Toyoda, A., Noguchi, H., et al. 2006, Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway, *Nat. Genet.*, **38**, 158–67.
25. Luo, Z.X., Ji, Q., Wible, J.R. and Yuan, C.X. 2003, An Early Cretaceous tribosphenic mammal and metatherian evolution, *Science*, **302**, 1934–40.
26. Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., et al. 2007, The delayed rise of present-day mammals, *Nature*, **446**, 507–12.
27. Margulies, E.H., Maduro, V.V., Thomas, P.J., et al. 2005, Comparative sequencing provides insights about the structure and conservation of marsupial and monotreme genomes, *Proc. Natl. Acad. Sci. USA*, **102**, 3354–9.
28. Yoder, J.A., Walsh, C.P. and Bestor, T.H. 1997, Cytosine methylation and the ecology of intragenomic parasites, *Trends Genet.*, **13**, 335–40.
29. Agrawal, A., Eastman, Q.M. and Schatz, D.G. 1998, Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system, *Nature*, **394**, 744–51.
30. Dupressoir, A., Vernochet, C., Bawa, O., et al. 2009, Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene, *Proc. Natl. Acad. Sci. USA*, **106**, 12127–32.
31. Mi, S., Lee, X., Li, X., et al. 2000, Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis, *Nature*, **403**, 785–9.
32. Nakamura, T.M. and Cech, T.R. 1998, Reversing time: origin of telomerase, *Cell*, **92**, 587–90.
33. Smit, A.F. 1999, Interspersed repeats and other mementos of transposable elements in mammalian genomes, *Curr. Opin. Genet. Dev.*, **9**, 657–63.
34. Freyer, C., Zeller, U. and Renfree, M.B. 2007, Placental function in two distantly related marsupials, *Placenta*, **28**, 249–57.
35. Freyer, C., Zeller, U. and Renfree, M.B. 2003, The marsupial placenta: a phylogenetic analysis, *J. Exp. Zool. A Comp. Exp. Biol.*, **299**, 59–77.
36. Freyer, C. and Renfree, M.B. 2009, The mammalian yolk sac placenta, *J. Exp. Zool. B Mol. Dev. Evol.*, **312**, 545–54.
37. Renfree, M.B. 2010, Marsupials: placental mammals with a difference, *Placenta*, **31**(Suppl. S21–26).