An endogenous retrovirus presumed to have been endogenized or relocated recently in a marsupial, the red-necked wallaby

Sakura Hayashi [1], Konami Shimizu [2], Yusuke Honda [2], Yukako Katsura [1], and Akihiko Koga [1]

[1] Primate Research Institute, Kyoto University, Inuyama City, Japan
[2] Noichi Zoological Park of Kochi Prefecture, Konan City, Japan

Keywords: albinism, marsupial, endogenous retrovirus, retrotransposon, long terminal repeat

Corresponding author: Akihiko Koga (email: koga.akihiko.5n@kyoto-u.ac.jp)

Abstract

An albino infant wallaby was born to a mother with the wild-type body color. PCR and sequencing analyses of *TYR* (encoding tyrosinase, which is essential for melanin biosynthesis) of this albino wallaby revealed a 7.1-kb-long DNA fragment inserted in the first exon. Because the fragment carried long terminal repeats, we assumed it to be a copy of an endogenous retrovirus, which we named *walb*. We cloned other *walb* copies residing in the genomes of this species and another wallaby species. The copies exhibited length variation, and the longest copy (>8.0 kb) contained open reading frames whose deduced amino acid sequences were well aligned with those of *gag*, *pol*, and *env* of retroviruses. It is not known through which of the following likely processes the *walb* copy was inserted into *TYR*: endogenization (infection of a germline cell by an exogenous virus), reinfection (infection by a virus produced from a previously endogenized provirus), or retrotransposition (intracellular relocation of a provirus). In any case, the insertion into *TYR* is considered to have been a recent event on an evolutionary timescale because albino mutant alleles generally do not persist for long because of their deleterious effects in wild circumstances.

Introduction

The lifecycle of retroviruses includes reverse transcription of their RNA genome and incorporation of DNA products into the chromosome of their host cells. If the incorporation occurs in a germline cell of an organism, the provirus is inherited by subsequent generations. This initial integration step is termed endogenization, and the provirus harbored since then, as part of the chromosome, is called an endogenous retrovirus (Boeke and Stoye 1997). An endogenous retrovirus may increase the per-cell copy number through two processes (Bannert and Kurth 2006). One is reinfection, that is, a virus particle produced from the provirus infects the same host individual and is endogenized. This may occur if the endogenous retrovirus copy retains all genetic information required for proliferation; however, even if the copy is in a defective form, its reinfection can occur through the provision of necessary factors by an exogenous virus. The other process is retrotransposition, that is, the provirus is transcribed and reverse transcribed, and the DNA produced is integrated into the chromosome of the same cell. The latter is basically the same as the amplification mechanism of long terminal repeat (LTR) retrotransposons. Based on this common mechanism and from similarities in structure, it is widely considered that endogenous retroviruses evolved from LTR retrotransposons and/or vice versa. Recent bioinformatic analyses across kingdoms provide support to the view that retroviruses were derived from LTR retrotransposons via acquisition of genes necessary for leaving the host cell and entering another (Eickbush and Jamburuthugoda 2008; Llorens et al. 2009).

Vertebrate genomes usually contain many families of endogenous retroviruses, where a family is defined as copies that are phylogenetically closely related, and each family consists of a large number of copies. For example, the total copy number of endogenous retroviruses in the human genome is ca. 80,000 (Lander et al. 2001; Paces et al. 2004). However, the vast majority of these copies are considered to have arrived at their current chromosomal positions in the remote past. The age of each provirus can be estimated from the difference in sequences between its LTRs; the LTR sequences are identical to each other when a provirus is newly formed; however, as time passes, differences occur and become larger owing to accumulation of mutations. Surveys of human endogenous retroviruses for the age distribution showed that the vast majority are as old as millions of years or more (Costas 2001; Bannert and Kurth 2006; Vargiu et al. 2016).

In our recent study, we clarified the genetic cause of an oculocutaneous albinism (red eyes and white coat) that occurred in an animal of the red-necked wallaby (*Macropus rufogriseus*). We found an extra 7.1-kb-long DNA fragment that was inserted in an exon of a protein-coding gene. The gene was *TYR* that encodes tyrosinase. Tyrosinase serves as an enzyme that catalyzes the first two reactions of melanin biosynthesis (Körner and Pawelek 1982). The extra fragment was subsequently revealed to be an endogenous retrovirus copy by

analyzing this and additionally cloned copies, one of which had the structure of a canonical retrovirus. We propose that, on an evolutionary timescale, the particular provirus found in *TYR* has recently been inserted. This proposal is based on two observations: its presence in the albinism-causing mutant gene against which there is strong natural selection, and identical sequences exhibited in its LTRs.

Materials and methods

Ethics

This study did not include any animal experiments; the materials used were a surplus of blood samples collected for another purpose and feces naturally defecated. This study involved a recombinant DNA experiment and it was approved in advance by the Recombinant DNA Experiment Safety Committee of Kyoto University (approval number 200421).

Animals and DNA preparation

In November 2015, an albino female wallaby (named Lily) was born to a mother of wild-type body color (named Ran) in Noichi Zoological Park (Konan City, Japan). Ran, born in February 2014, was legally imported in April 2015 from a kangaroo farm in the Netherlands and was subsequently found to have already been pregnant at the time of arrival. In November 2020, we extracted genomic DNA from a surplus of Lily's blood samples that were collected during emergency surgery. For comparison, in December 2020, we prepared genomic DNA from the feces of a wild-type animal that was born and raised in the same zoo. Hereafter, in this report, the wild-type and albino animals (Fig. 1), as well as their DNA samples, are denoted as WT and AL, respectively.

In July 2021, we extracted, genomic DNA from the feces of a tammar wallaby (*Macropus eugenii*). This animal (named Noma) was an adult female kept at Kuragaike Park Zoo (Toyota City, Japan) and exhibited the standard body color of this species.

PCR, cloning and sequencing

In most vertebrates hitherto examined, the *TYR* gene consists of five exons (denoted as Ex1 to Ex5 hereafter). We conducted BLAST searches of the genome databases of tammar wallaby (Meug_1.0 assembly) and koala (phaCin_unsw_v4.1 assembly), using the human *TYR* exon sequences (Ensembl ENST00000263321.6) as queries. We drew sequences that corresponded to the exons along with their flanking regions and selected 30-nucleotide blocks that encompassed the entire exon region. We then prepared PCR primers that represented these blocks (see Table S1 for sequences and origins). PCR amplification was performed under the same conditions as those previously

described (Koga et al. 2020; Mae et al. 2020). PCR products were cloned into the pUC118 vector and sequenced by the Sanger method. Sequencing primers were prepared step by step based on the sequence data obtained.

Results

PCR amplification of *TYR* exons

All five *TYR* exon regions were amplified (Fig. 2). In the assay for Ex1, while WT yielded a fragment of the expected length (1.5 kb), two fragments of larger sizes (8.6 kb and 1.9 kb) were observed in the lane for AL. There was no detectable size difference between WT and AL in the products for Ex2 to Ex5, and all these fragments exhibited the expected lengths.

Structure of Ex1 PCR products

We cloned the single product from WT (denoted as WT1.5) and the two fragments from AL (AL8.6 and AL1.9) separately into the pUC118 vector, and sequenced their entire insert portions (Fig. 3, Fig. S1).

Human *TYR* Ex1 contains an 819-bp segment that encodes the first 273 amino acids of the tyrosinase protein. The WT1.5 fragment contained an 819-bp sequence that was aligned with no gaps to the human coding sequence. The reduced amino acid sequence obtained from this 819-bp portion did not contain a stop codon and exhibited an identity as high as 82% (224/273). Based on these results, we regarded WT1.5 as a fragment that included Ex1 of the wallaby *TYR* gene.

The AL8.6 fragment carried a 510-bp sequence that corresponded to the first 510 bp of Ex1, an extra segment of 7137 bp, and a 315-bp sequence that corresponded to the last 315 bp of Ex1. The sum of the lengths of the two separate Ex1 regions (510 + 315 = 825 bp) was 6 bp longer than the length of Ex1 contained in WT1.5 (819 bp). This was because of the presence of the last six nucleotides of the first Ex1 part (GACATA) at the head of the second Ex1 part, which is reminiscent of a target site duplication created upon integration. The extra 7137-bp fragment (registered at GenBank with accession number LC631488) contained five open reading frames (ORFs) for 709-, 160-, 218-, 783-, and 315-amino-acid peptides. It was also found that the first and the last 389-bp sequences of the extra 7137-bp fragment were identical to each other, which is reminiscent of an endogenous retrovirus.

The AL1.9 fragment contained the first 510-bp portion of Ex1, a 389-bp sequence identical to the 389-bp terminal regions of the insertion in AL8.6, and the last 315-bp portion of Ex1. A 6-bp duplication of GACATA was also found.

Properties of insertion fragments

A BLAST search using the 7137-bp inserted fragment as a query sequence

returned a high-score hit to TvERV (Baillie and Wilkins 2020) which is a type-D endogenous retrovirus of the common brushtail possum *Trichosurus vulpecula*, an Australian marsupial. The five ORFs found in the 7137-bp fragment shared a high similarity in reduced amino acid sequence with the genes contained in TvERV. Detailed results are described later as part of the results of the structural analysis of this and additional copies. The possession of ORFs featuring an endogenous retrovirus, along with the repeat structures at and outside the ends, indicates that the 7137-bp fragment is an endogenous retrovirus. It was possible that this particular insertion sequence might be one of multiple copies that descended from a common ancestor. Assuming such a situation, we named the putative multicopy family *walb* and the particular 7137-bp copy found in the red-necked wallaby *walb*_red01.

A solitary LTR is known to be frequently generated in the host genome by homologous recombination between two LTRs (Vitte and Panaud 2003). The AL1.9 fragment contained a single LTR in *TYR* Ex1, which could be explained as a solitary LTR.

Origin of AL1.9

A possible interpretation of the Ex1 PCR results is that the *TYR* locus of the AL animal is heterozygous; one *TYR* allele carries the entire *walb* element and the other allele contains a solitary LTR. However, it needs to be considered that PCR amplification frequently results in the generation of an artifact product when the template DNA contains a repeat structure. In the case of template DNA with LTRs, a strand synthesized without sufficient elongation can be annealed to the other LTR in the next cycle and serve as a primer. Once a short fragment is generated, it is expected to have a higher amplification efficiency because of its shorter length. To evaluate this possibility, we performed a PCR assay with template DNA that mimicked diploid genomic DNAs of *TYR* homozygotes or *TYR* heterozygote (Fig. S2). We demonstrated that (i) an AL1.9 fragment is produced from the AL8.6 clone as a PCR artifact, (ii) amplification efficiency in AL1.9 is higher than that in AL8.6 to the extent that AL8.6 is not detectable in the heterozygote mimicry (Hum+LS), and (iii) the quantitative ratio of the two fragments in the product from the homozygote mimicry (Hum+LL) was similar to that from the AL genomic DNA (AL lane in Fig. 3). These results indicate that the AL animal is homozygous for the allele that corresponds to AL8.6.

Search of other marsupial genomes for *walb*

A BLAST search against the tammar wallaby genome database (Meug_1.0), using the entire *walb*_red01 sequence (7137 bp) as a query, returned a high-score hit to one sequence, and the entire *walb* LTR sequence (389 bp) was aligned to a 388-bp sequence with one gap and a nucleotide identity of 95% (371/388). The sequence alignment is shown later, together with the results of structural analysis. The lack of a hit to the *walb* internal region could be

interpreted as the presence of a solitary LTR in tammar wallaby, but this sequence may be an artifact generated by a sequencing or assembly error during the database construction. We also conducted the same searches against the latest database versions of koala, wombat, Tasmanian devil, and opossum; however, a nearly complete similarity like that observed with tammar wallaby was not found.

PCR to detect LTR-encompassed regions

To determine whether the internal region of *walb* is present in the tammar wallaby genome, and to find other *walb* copies in the red-necked wallaby, we conducted a PCR assay by devising primers suitable for this purpose. A PCR assay to amplify a sequence that includes two LTRs causes an artifact product, as revealed with AL1.9. To avoid this, we prepared a forward primer that represented a nucleotide block in the 3' half of the LTR, and a reverse primer in the 5' half of the LTR (Fig. 4A). A PCR assay using these primers yielded a single fragment of approximately 2.6 kb from genomic DNA of the red-necked wallaby, and several fragments of 5 to 9 kb from the tammar wallaby (Fig. 4B). These results suggested that the red-necked wallaby contains a *walb* copy of approximately 2.6 kb, and that the tammar wallaby possesses *walb* copies of various lengths. We combined the PCR products with the plasmid vector, introduced the DNA into bacteria, and picked up individual bacterial colonies. Seven of the eight red-necked wallaby clones we examined were found to carry an insert fragment and all of these insert fragments were 2.6 kb in length. One of these was named *walb*_red02 and sequenced for its structure (registered in GenBank with accession number LC647198). Twelve of the 16 tammar wallaby clones we checked were found to have insert fragments of different lengths. Of these 12 clones, four carried insert fragments of >7 kb (Fig. 4C). In the electrophoresis image shown in Fig. 4C, the four clones were arranged in order of length and designated as *walb*_tam01 to *walb*_tam04. We were interested in finding *walb* copies which were longer than *walb*_red01 (7137 bp) because longer copies might have a structure closer to that of the original retrovirus. We sequenced the entire insert portions of the two longest clones (*walb*_tam01 and _tam02; registered in GenBank with accession numbers LC647196 and LC647197, respectively).

Structure of the internal region of the *walb* copies

The three *walb* copies of >7 kb in length (*walb*_red01, _tam01, and _tam02) consisted of the same components (Fig. 5). Pairwise comparisons revealed that many of the components of *walb*_red01 and *wal*_tam02 were slightly shorter than those of *walb*_tam01, and that these shorter lengths result from the lack of small nucleotide blocks. This structural relationship suggested that *walb*_tam01 is a copy identical or close to a canonical endogenous retrovirus and the other three copies are variant sequences derived through partial deletions. The

*walb*_red02 was a copy in which an internal region as large as 5.5 kb was deleted.

Three ORFs were identified in *walb*_tam01. Their deduced amino acid sequences were aligned, without large deletions or insertions, to those of the genes contained in TvERV (Fig. 6). The first ORF corresponded to TvERV *gag*, the second ORF to *pro* and *pol*, and the third ORF to *env*. Based on these clear correspondences, we denote the three ORFs as *gag*, *pro*/*pol*, and *env* in Fig. 5 and Fig. 6. Similar to TvERV and many other LTR retroelements, the head region of *pro* was overlapped with the tail region of *gag*, and there was a frameshift between them. A typical Lys-tRNA primer binding site was identified upstream of *gag*, and a typical polypurine tract was found downstream of *env*.

The source virus of TvERV is considered to belong to the *Betaretrovirus* genus (Escalera-Zamudio et al. 2015). Assuming that this is also the case with *walb*, we constructed a phylogenetic tree of several known betaretroviruses (Fig. 7) plus *walb*. A gammaretrovirus (KoRV) was also added for comparison. TvERV and *walb* formed a cluster in this tree.

Structure of LTR

The *walb*_red01 clone contained the entire LTR sequence because the primers used for PCR amplification were those located outside the element. In contrast, in the other three clones, the 5' terminal region of the 5' LTR and the 3' terminal region of the 3' LTR remained unknown because the PCR primers used were located inside the LTRs. To overcome this and draw available information, we analyzed the structure of the 388-bp sequence that was aligned to the *walb*_red01 LTR (389 bp) and identified in the tammar wallaby genome database. The CAAT box, TATA box, poly(A) addition signal, and terminal inverted repeats are standard components of retrovirus LTRs. These were all found in the two LTR sequences examined (Fig. 8).

Discussion

Cause of the albino phenotype

In AL8.6, the first 510-bp sequence of Ex1 was followed by the *walb*_red01 sequence. It is an acceptable assumption that during transcription of this *walb*-carrying *TYR* gene, RNA polymerase moves along the template strand and enters the *walb* portion. A peptide produced by translation of mRNA generated in this manner is expected to comprise 226 amino acids because a termination codon (TAA) occurs at the 227th site. In this peptide, the sequence of the first 170 amino acids of tyrosinase is represented, but an unrelated sequence comprising 56 amino acids follows it. Human tyrosinase consists of 530 amino acids. If we assume that the wallaby tyrosinase is similar in size, a peptide that lacks approximately the last two-thirds portion is produced from the *walb*-

carrying *TYR* gene of the AL animal. This is highly probable because a typical splicing donor site (MAGGTRAGT, using the IUPAC nucleotide codes) (Zhang et al. 2003), or even its core nucleotide block (AGGT), was not found before the stop codon.

The *walb*_red01 LTR contains promoter signals (Fig. 8), and a transcript may be produced through the function of these signals. A peptide produced from this transcript lacks approximately the first one-third portion of the wallaby tyrosinase.

There are many reports of human albinism due to mutations in *TYR* (the P14679 file of the UniProtKB database). These reports include examples of oculocutaneous albinism caused by mutational amino acid changes at sites after the 170th site, including V177F, P205T, and C289G. There are also examples of albinism due to changes located before the 170th site, including H19Q, P81L, and G109R. These examples suggest that the *walb* insertion in *TYR* is the cause of the albino phenotype of the AL animal. In addition to the insertion, the AL *TYR* may carry other mutation(s) that affect tyrosinase function. However, the *walb* insertion is considered to be so critical that the insertion alone can cause a tyrosinase malfunction.

Origin of the mutant *TYR* gene

The AL animal (Lily) was homozygous for the *walb*-carrying allele. Her mother (Ran) exhibited the wild-type body color, suggesting that Ran was heterozygous and the *walb*-carrying allele is recessive to the wild-type allele. Because Ran is already dead and no tissue sample is available, we cannot confirm this directly. Lily's father has not been identified but is likely to have the wild-type body color. Both parents were born in a kangaroo farm in the Netherlands, which suggests that Lily's inbreeding coefficient may be high. The *walb*-carrying allele is thought to have already been present in a founder animal of the farm. We do not have information to estimate how long the mutant allele was maintained until reaching this particular founder animal; however, based on the reasoning described below, the time span is considered to be short on the evolutionary scale.

Properties as an endogenous retrovirus

The *walb*_tam01 copy was the longest of the four *walb* copies that we cloned and analyzed. It contained three ORFs that corresponded to *gag*, *pol*/*pro*, and *env* of retroviruses, and other signaling motifs commonly found in retroviruses. Although the entire LTR regions of *walb*_tam01 could not be obtained because of the limitation in probe design, one can expect that *walb*_tam01 carries LTRs similar in sequence to those of *walb*_red01. Our survey of the *walb*_red01 LTRs revealed various motifs that are usually found in retroviruses. These structural features led to our conclusion that *walb* is an endogenous retrovirus, which was initially only an assumption in our study. Our phylogenetic analysis suggested a

close relationship between the retrovirus from which *walb* originated and betaretroviruses.

Cause of structural variation

The *walb*_red01, _red02, and _tam02 copies were shorter in length than *walb*_tam01, mainly because of internal deletions. It is known that errors in template switching during reverse transcription are common and these frequently result in internal deletions (Bruner et al. 2016). The most likely cause of the emergence of the three shorter copies would be the occurrence of such errors at the time of endogenization. It is another possibility that internal deletions occurred in the DNA sequences while they were inherited as proviruses. These are not mutually exclusive; the shorter copies may have experienced multiple sequence alteration events.

Origin of *walb*

Our PCR experiment (Fig. 5) and subsequent cloning and sequencing analysis (Fig. 6) showed that both the red-necked wallaby and tammar wallaby harbor multiple *walb* copies in their genomes and these copies have sequence variations. This situation may have resulted from a *walb* endogenization which occurred before the divergence of the two wallaby species, that is, in their common ancestor. Another possibility is that *walb* was endogenized in each species after the species split. These two scenarios are not mutually exclusive; endogenization may have occurred both before and after the speciation.

How widely *walb* is distributed in the phylogeny of marsupials, or even mammals, is not known. Although our BLAST searches did not find a *walb*-like sequence in the available genome databases, including those of koala and opossum, this does not necessarily imply that *walb* is absent in their genomes. It should be noted that the BLAST search against the tammar wallaby database failed to find a *walb* copy, except for a single LTR sequence. Multiple *walb* copies do exist in the tammar wallaby genome, as demonstrated in our PCR and sequencing experiments. This discrepancy may have been caused by the tendency of repetitive sequences to be underrepresented in genome databases (Koga 2012).

Recent insertion

Considering that *walb* is an endogenous retrovirus, we can propose the following three hypotheses concerning the integration of *walb*_red01 into *TYR*: (i) endogenization: an individual of the red-necked wallaby was infected by a virus that was produced in another individual or in another host species, and the virus DNA was inserted into the *TYR* of a germline cell; (ii) reinfection: a virus produced from a previously endogenized provirus infected the same host individual and DNA was inserted into the *TYR* of a germline cell; and (iii) retrotransposition: a previously endogenized provirus was transcribed and

reverse transcribed, and the DNA product was inserted into *TYR*, all occurring within a single germline cell. The second hypothesis involves reinfection solely by the provirus and reinfection with participation of an exogenous virus. We do not have sufficient information to consider which of these is most likely to have occurred. In any of these cases, however, it can be inferred that this germline insertion event occurred recently on an evolutionary timescale based on two observations.

One is the linkage between *walb*_red01 and the mutant *TYR* allele, the latter causing oculocutaneous albinism. Mutant animals with oculocutaneous albinism generally have serious disadvantages in terms of survival in wild circumstances, because of a higher chance of detection by predators, higher chance of escape of prey, decrease in UV protection, and decrease in visual acuity. These disadvantages result in the exclusion of albino mutant alleles from the gene pool of the host organism. Thus, even if an albino allele emerges through mutation, it does not persist long in the population of the host organism. This implies that when an existing albino allele is found, its origin is expected to be recent. If the mutant allele carries an insertion sequence, the time span from the occurrence of insertion is expected to be short. Although albino mutant alleles are mostly recessive to wild-type alleles, the recessive nature does not invalidate the above explanation. Suppose that a recessive allele emerges, and that strong natural selection acts against its homozygous carrier. If the genotype frequency of the heterozygote is increased, the frequency is driven back by the selection pressure to eliminate the mutant homozygote. When the genotype frequency of the heterozygote is low, the mutant allele is constantly exposed to a chance of stochastic loss.

Another observation that supports the inference of recent occurrence is that the LTRs of *walb*_red01, both 389 bp in length, were identical in sequence. In relation to this, analysis of LTRs of additional *walb* copies may provide useful information on whether the insertion of *walb*_red01 into *TYR* occurred uniquely in the genome or it was one of multiple events that occurred simultaneously or as a series of events.

Author contributions

A.K. conceived of the study and designed the experiments. K.S. and Y.H. conducted animal care and sample collection. A.K. and S.H. carried out the experimental work. S.H. and Y.K. conducted data analysis.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We are grateful to Kuragaike Park Zoo (Toyota City, Japan) for providing fecal

samples of the tammar wallaby. We also than Yuki Enomoto (Kyoto University) for technical assistance. This work was supported by Grants-in-Aid from the Japan Society for the Promotion of Science (grant number 19H03311 to A.K.).

Supplementary material

This article contains electronic supplementary material (Table S1, Fig. S1, Fig. S2).

References

Baillie, G.J., and Wilkins, R.J. 2001. Endogenous type D retrovirus in a marsupial, the common brushtail possum (*Trichosurus vulpecula*). J. Virol. 75, 2499-2507. (doi: 10.1128/JVI.75.5.2499-2507.2001)

Bannert, N,, and Kurth, R. 2006. The evolutionary dynamics of human endogenous retroviral families. Annu. Rev. Genomics Hum. Genet. **7**, 149-173. (doi: 10.1146/annurev.genom.7.080505.115700)

Boeke, J.D., and Stoye, J.P. 1997. Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In: Coffin, J.M., Hughes, S.H., and Varmus, H.E., editors. Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press

Bruner, K.M., Murray, A.J., Pollack, R.A. et al. 2016. Defective proviruses rapidly accumulate during acute HIV-1 infection. Nat. Med. **22**, 1043-1049. (doi: 10.1038/nm.4156)

Costas, J., 2001. Evolutionary dynamics of the human endogenous retrovirus family HERV-K inferred from full-length proviral genomes. J Mol Evol. 2001 Sep;53(3):237-43. doi: 10.1007/s002390010213. PMID: 11523010.

Eickbush, T.H., and Jamburuthugoda, V.K. 2007. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res. **134**, 221-234. (doi:10.1016/j.virusres.2007.12.010)

Koga, A. 2012. Under-representation of repetitive sequences in whole-genome shotgun sequence databases: an illustration using a recently acquired transposable element. Genome **55**, 172-175. (doi: 10.1139/g11-088)

Körner, A., and Pawelek, J. 1982. Mammalian tyrosinase catalyzes three reactions in the biosynthesis of melanin. Science **217** (4565), 1163-1165.

Kumar, S,, Stecher, G., Li, M., Knyaz, C., and Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol. Biol. Evol. **35**, 1547-1549. (doi: 10.1093/molbev/msy096)

Lander, E.S., Linton, L.M., Birren, B. et al. (International Human Genome Sequencing Consortium). 2001. Initial sequencing and analysis of the human genome. Nature **409**, 860–921. (doi: 10.1038/35057062)

Llorens, C., Muñoz-Pomer, A., Bernad, L. et al. 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. Biol. Direct **4**, 41. (doi.org/10.1186/1745-6150-4-41)

Paces, J., Pavlícek, A., Zika, R., Kapitonov, V.V., Jurka, J., and Paces, V. 2004. HERVd: the Human Endogenous RetroViruses Database: update. Nucleic Acids Res. **32**, D50. (doi: 10.1093/nar/gkh075)

Vargiu, L., Rodriguez-Tomé, P., Sperber, G.O., et al. 2016. Classification and characterization of human endogenous retroviruses; mosaic forms are common. Retrovirology 22, 7. (doi: 10.1186/s12977-015-0232-y)

Vitte, C., and Panaud, O. 2003. Formation of Solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR

retrotransposons in rice *Oryza sativa* L. Mol. Biol. Evol. **20**, 528-540. (doi: 10.1093/molbev/msg055)

Zhang, X.H., Heller, K.A., Hefter, I., Leslie, C.S., and Chasin, L.A. 2003. Sequence information for the splicing of human pre-mRNA identified by support vector machine classification. Genome Res. **13**, 2637–2650. (https://doi.org/10.1101/gr.1679003)

Fig. 1. Wallaby animals. The wild-type (WT) animal has black eyes and a body color of the regular pigmentation pattern of the red-necked wallaby. The albino (AL) animal exhibits oculocutaneous albinism, having red eyes, white hair, and pinkish skin. Their body height is approximately 80 cm.

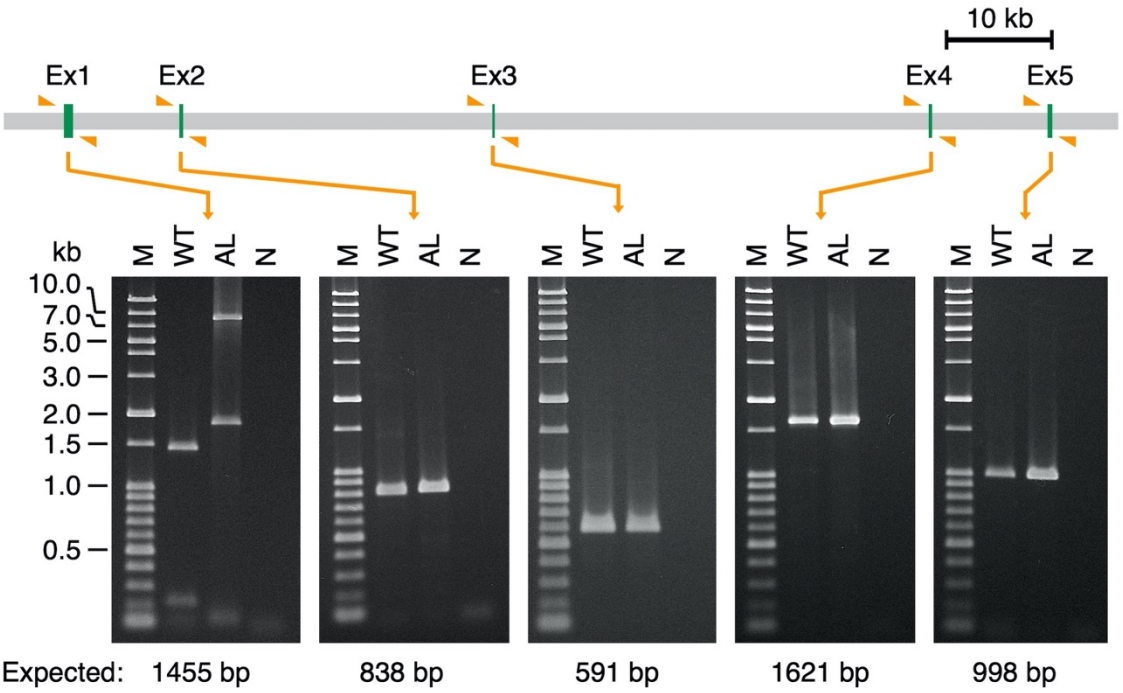


Fig. 2. Results of PCR amplification for *TYR* exon regions. The structure of human *TYR* is shown at the top. Green squares are the exons (Ex1 to Ex5) and the grey bar indicates introns or untranscribed regions. Orange triangles indicate the location and direction of PCR primers to amplify each exon region. The elements do not strictly reflect their sizes or positions. Electrophoresis photographs of PCR reaction mixtures are shown below. The lanes are: M, size marker; WT, WT template; AL, AL template, and N, no template DNA. The expected product size based on the tammar wallaby or koala sequence information is shown below each photograph.

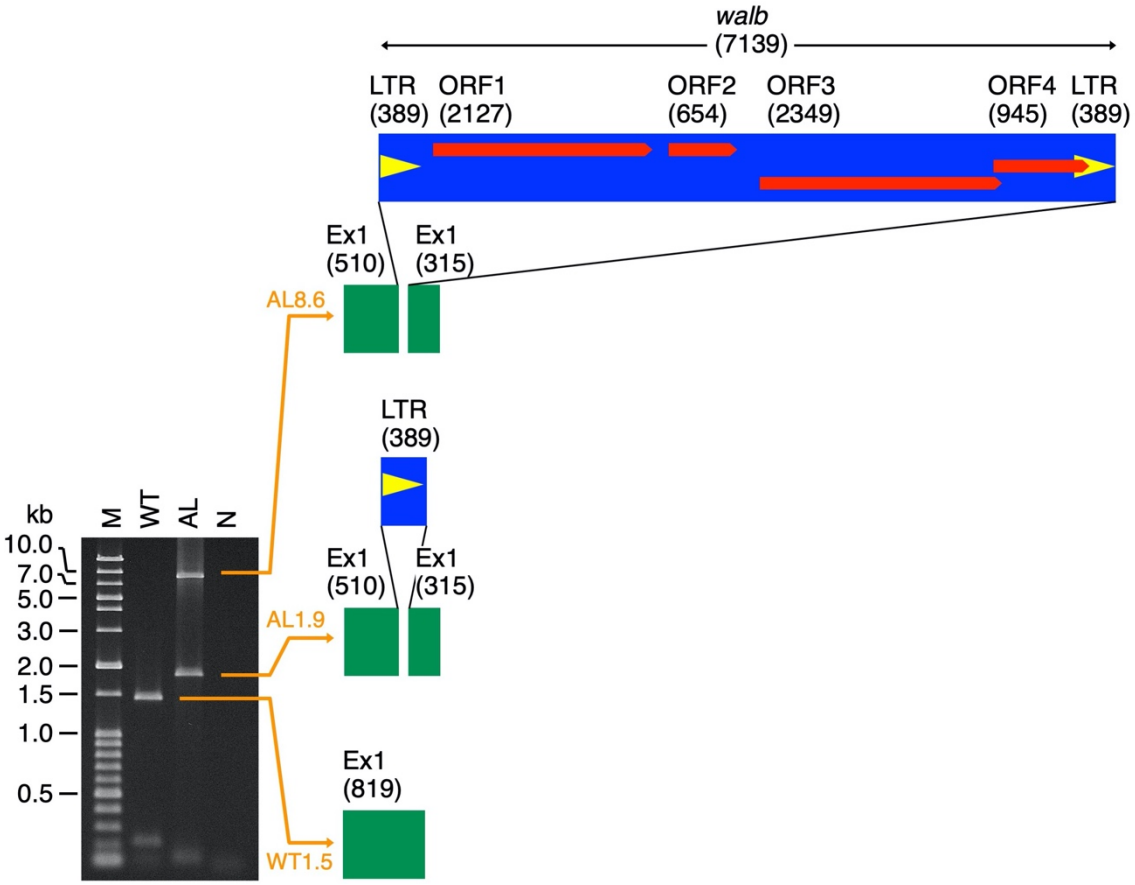


Fig. 3. Structure of Ex1 PCR products. The photograph of Ex1 PCR products is shown here again. Each product fragment was cloned and sequenced, and analyzed for structure. The size in bp of each component is shown in parentheses.

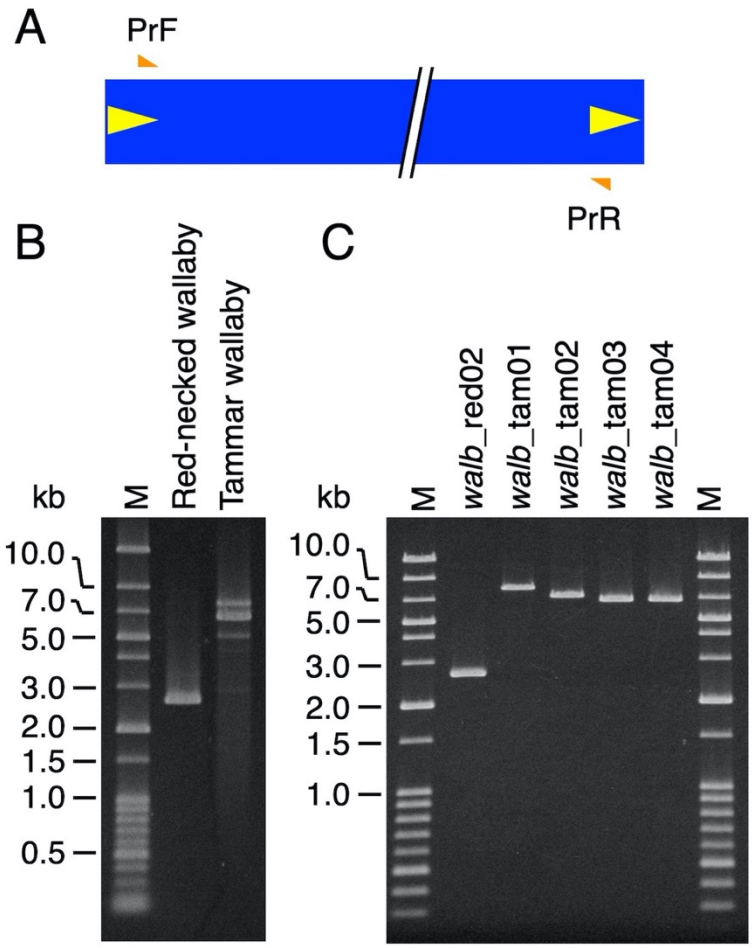


Fig. 4. PCR to amplify LTR-encompassed regions. **A**. Locations and directions of the PCR primers. The yellow triangles indicate the LTRs. PrF and PrR represent nucleotide blocks in the LTR sequence as shown here. The exact locations of PrF and PrR were nt 202 to 236 and nt 150 to 116, respectively, on the *walb* sequence (GenBank LC631488) in which the 5' LTR consisted of nt 1 to 389. **B**. PCR from genomic DNA. The template DNA was genomic DNA of the two wallaby species. **C**. PCR from *walb* clones. The template DNA was the *walb* clone that originate from the red-necked wallaby, and the four clones from the tammar wallaby that exhibited fragment lengths of >7 kb.

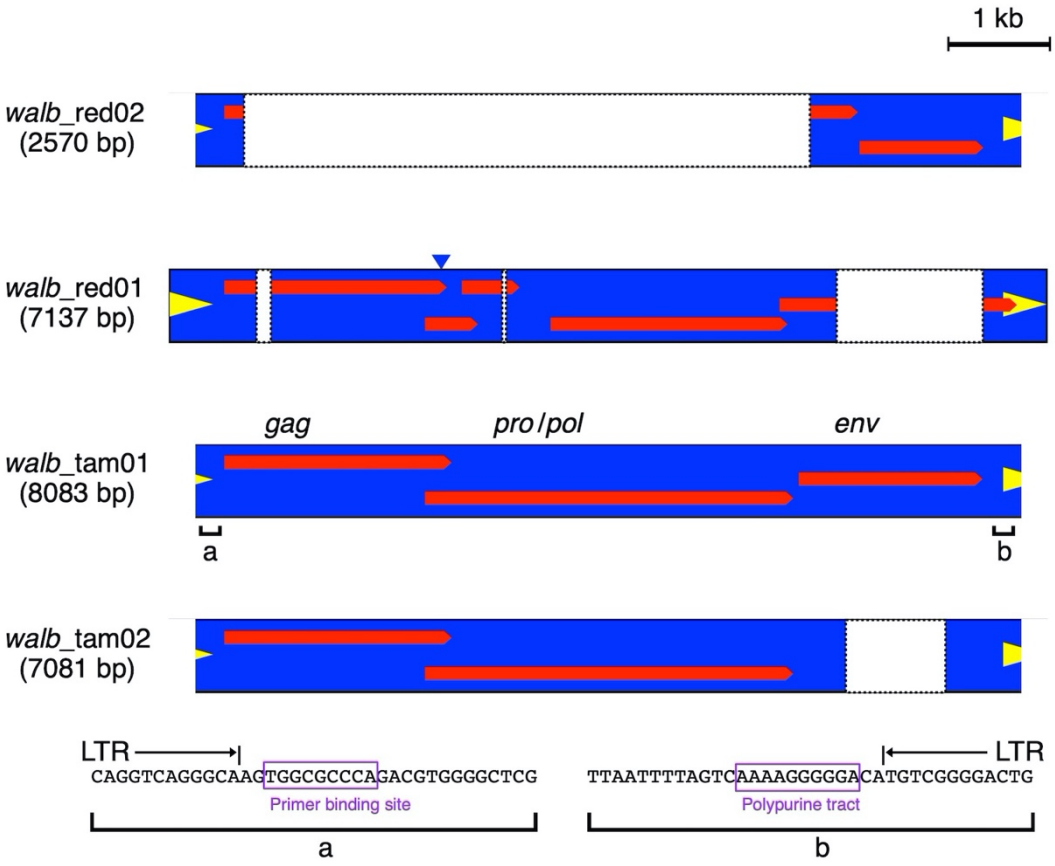


Fig. 5. Structure of the *walb* clones. The *walb*_red01 copy was cloned by PCR using primers that represented *TYR* (see Fig. 2). Because these primers were located outside of the *walb* element, the clone obtained included the two entire LTRs. The PCR primers for the other three clones were located inside the LTRs (see Fig. 4) and therefore, the *walb* terminal regions were not included in the clones obtained. The longest clone was *walb*_tam01, and the other three clones lacked one or more portions when compared with *walb*_tam01. The deleted regions are shown by dotted boxes. The blue triangle on *walb*_red01 indicates an extra nucleotide in comparison to *walb*_tam01. The ORFs represented by red arrows are those identified in the sequence of each clone. Because the ranges of ORFs are affected by deletions or other sequence variation, the identity of the ORFs among the clones was not taken into consideration. The sequences of regions a and b regions are shown at the bottom, with sequence motifs present therein.

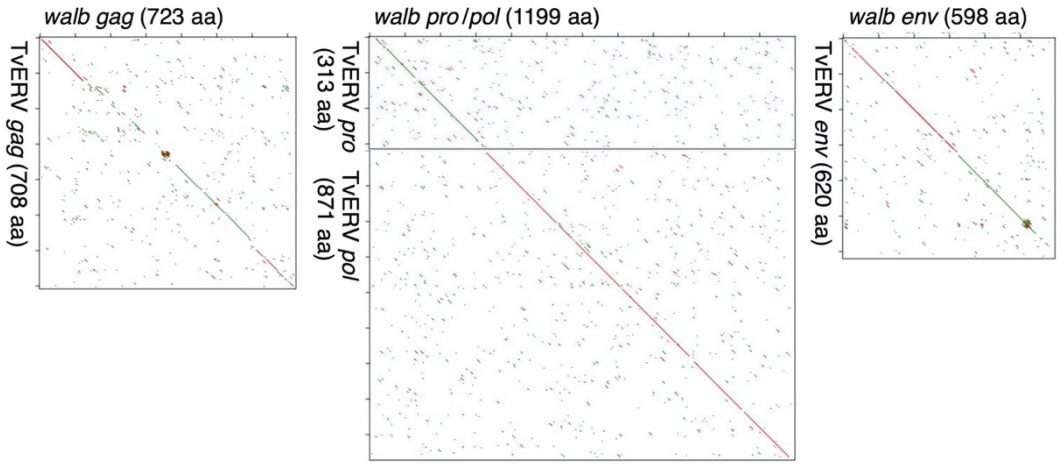
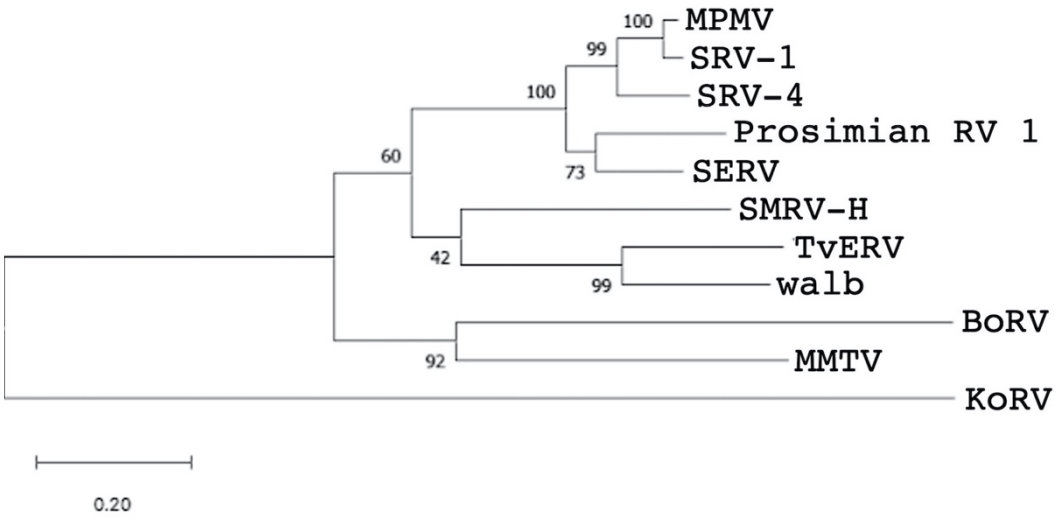


Fig. 6. Dot matrix analysis of amino acid sequences. Reduced amino acid sequences from the *walb*_tam01 ORF nucleotide sequences (horizontal axis) were compared with those of *gag*, *pro*, *pol* and *env* of TvERV (vertical axis). The criteria were a match of three or more amino acids over a window of 10 amino acids. The axes are marked with an interval of 100 amino acids.



Virus	Host organism	GenBank
TvERV	brushtail	AF224725
MMTV	mouse	AF228552
MPMV	rhesus macaque	AF033815
SERV	baboon	U85505
SRV-1	rhesus macaque	M11841
SMRV-H	human	M23385
BoRV	bovine	NC_029852
Prosimian RV 1	lemur	MT787217
SRV-4	crab-eating macaque	NC_014474
KoRV	koala	NC_039228
walb	wallaby	LC647196

Fig. 7. Phylogenetic tree of betaretroviruses, a gammavirus, and *walb*. Using the MEGA X platform (Kumar et al. 2018), amino acid sequences corresponding to the *pol* gene were aligned by the MUSCLE program and their phylogenetic tree was constructed by the maximum likelihood method. The bootstrap values were obtained through 1000 repetitions. The table under the tree shows the host organism of the viruses analyzed and their GenBank accession numbers. KoRV is a gammavirus, which was included in the analysis for comparison.

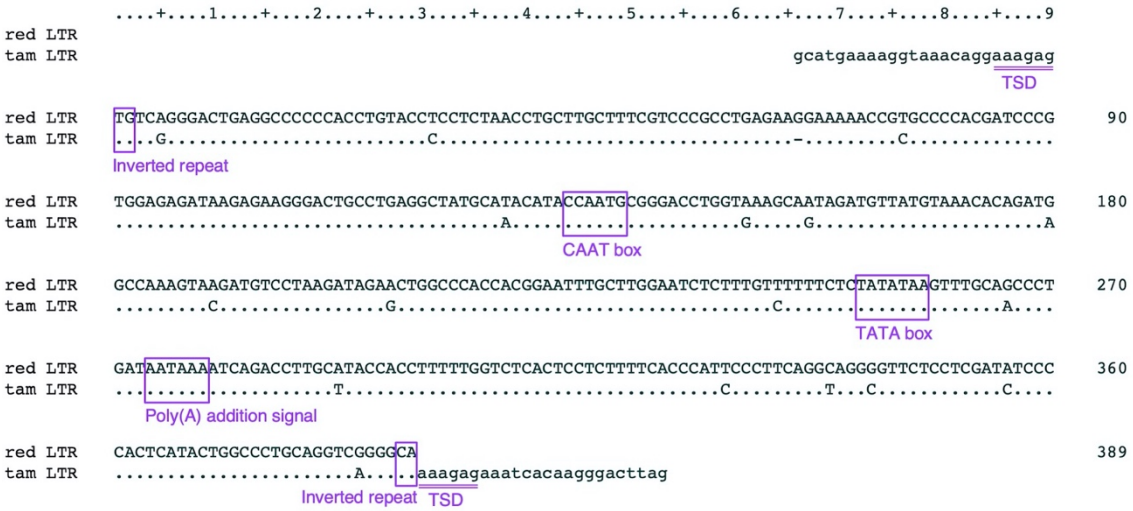


Fig. 8. Alignment of LTR sequences and motifs found therein. The upper line (labelled red LTR) is the sequence of the LTR that was contained in the *walb*_red01 clone. The lower line (tam LTR) is the sequence identified by a BLAST search of the tammar wallaby genome database. The dot indicates the occupation of the site by an identical nucleotide. The minus symbol implies absence of a site at its position. The motifs found are shown by purple boxes. The underline indicates 6-bp blocks that are considered to be target site duplications.