VIRUS EVOLUTION

Virus Evolution, 2022, **8(2)**, 1–7
DOI: https://doi.org/10.1093/ve/veac084
Advance Access Publication date: 3 September 2022
Rapid Communication

Monotreme-specific conserved putative proteins derived from retroviral reverse transcriptase

Koichi Kitao,[1] Takayuki Miyazawa,[1,†,*] and So Nakagawa[2,‡,*]

[1]Laboratory of Virus-Host Coevolution, Institute for Life and Medical Sciences, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan and
[2]Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan
[†]https://orcid.org/0000-0003-1938-9661
[‡]https://orcid.org/0000-0003-1760-3839
*Corresponding authors: E-mail: takavet@infront.kyoto-u.ac.jp; so@tokai.ac.jp

Abstract

Endogenous retroviruses (ERVs) have played an essential role in the evolution of mammals. ERV-derived genes are reported in the therians, many of which are involved in placental development; however, the contribution of the ERV-derived genes in monotremes, which are oviparous mammals, remains to be uncovered. Here, we conducted a comprehensive search for possible ERV-derived genes in platypus and echidna genomes and identified three reverse transcriptase-like genes named *RTOM1*, *RTOM2*, and *RTOM3* clustered in the *GRIP2* intron. Comparative genomic analyses revealed that *RTOM1*, *RTOM2*, and *RTOM3* are strongly conserved and are under purifying selection between these species. These could be generated by tandem duplications before the divergence of platypus and echidna. All *RTOM* transcripts were specifically expressed in the testis, possibly suggesting their physiological importance. This is the first study reporting monotreme-specific *de novo* gene candidates derived from ERVs, which provides new insights into the unique evolution of monotremes.

Key words: endogenous retrovirus; monotreme; reverse transcriptase; virus-derived gene.

Introduction

Endogenous retroviruses (ERVs) are remnants of retroviral genomes found in the host genomes. ERVs are retroviruses that infected the host germline cells and were integrated into the host genome (Johnson 2019). Young ERVs retain their viral open reading frames (ORFs) but gradually lose their intact ORFs due to the accumulation of mutations. However, proteins expressed from ERVs sometimes evolve as functional genes in the host (Ueda et al. 2020). A typical example is the syncytin genes, ERV-derived fusogenic genes, which are expressed in the human placenta (Blaise et al. 2003; Blond et al. 2000; Mi et al. 2000) and are essentially required for mouse placenta formation (Dupressoir et al. 2009, 2011). Syncytin genes have been independently acquired from different ERVs in different mammalian lineages, which is a representative example of the convergent evolution (Imakawa, Nakagawa, and Miyazawa 2015). In addition, other ERV-derived genes that do not show fusogenic activity have also been found to be expressed in the placenta. For example, *HEMO* encoding a secreted envelope protein (Heidmann et al. 2017) as well as *gagV1* and *pre-gagV1* genes (Boso et al. 2021) are highly expressed in the human placenta. Restriction factors against exogenous retroviruses are another example of viral gene co-option. For example, *gag*-derived *Fv1* (Best et al. 1996) and *env*-derived *Fv4* (Ikeda and Sugimura 1989) inhibit retroviral infection in mice. Despite these contributions to the evolution of therians, it is still unclear whether ERV-derived genes are co-opted in monotremes (egg-laying mammals).

Here, we attempted to determine whether there are ERV-derived genes specific to monotremes. Comparative studies for the detection of ERV-derived genes have been conducted in mammalian genomes, including the platypus (Nakagawa and Takahashi 2016; Wang and Han 2020). However, for monotremes, only the genome sequence of one species, the platypus, was available (OANA5), the quality of which was limited (Warren et al. 2008). Recently, high-quality monotreme genomes of platypus (mOrnAna1.p.v1) and echidna (mTacAcu1.pri) were determined using long-read sequencing technology (Zhou et al. 2021). Taking advantage of these genome sequences, we conducted comparative analyses and detected three novel ERV-derived genes specific to the monotreme lineage.

Results and discussion

To comprehensively search for ERV-derived genes in the genomes of monotremes, we extracted ORFs from the genomes of platypus and echidna. The amino acid sequences obtained by the virtual translation of these ORFs were used as queries for the sequence search. We used the hidden Markov model (HMM) of retroviral genes in the Gypsy Database 2.0 (GyDB) (Llorens et al. 2011) as

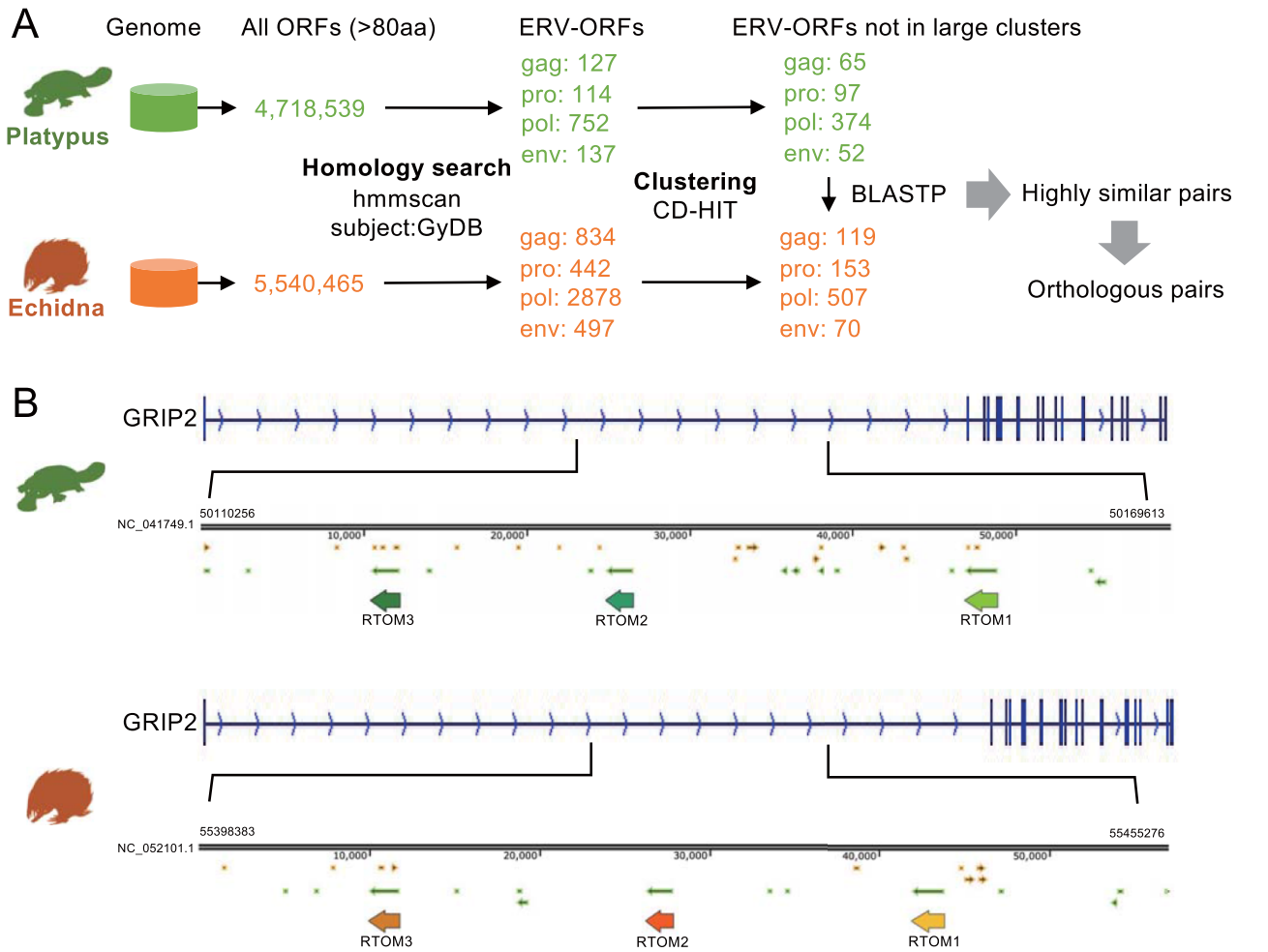


Figure 1. Identification of *RTOM1*, *RTOM2*, and *RTOM3*. (A) Schematic representation of the in silico screening for conserved ERV-derived genes in platypus and echidna. (B) Genomic context of *RTOM1*, *RTOM2*, and *RTOM3*. The thin arrows indicate ORFs above 100 amino acid length. The nucleotide sequences of *RTOM* ORFs are available in Supplementary File 2.

the subject of the sequence search (Supplementary Table S1). We identified ORFs similar to *gag*, *pro*, *pol*, and *env* genes (Fig. 1A). These ORFs are presumed to be a mixture of ORFs that (1) have physiological functions and are evolutionarily conserved and (2) have not been disrupted by nonsense mutations by chance. Most ERV ORFs belonging to (2) are derived from young ERVs, in which mutations have not yet accumulated. To exclude young ERV ORFs, we performed the clustering analysis based on the amino acid sequence identity. Since young ERVs are thought to be included in large clusters due to their mutual similarity, we removed sequences that belonged to large clusters consisting of more than ten sequences. This step could also exclude evolutionarily conserved but highly duplicated genes such as SCAN domain-containing genes (Emerson and Thomas 2011), which is beyond the scope of this study. Next, using the platypus ORFs as queries and the echidna ORFs as the subjects, we conducted a sequence similarity search using BLASTp. We obtained ORF pairs with high amino acid similarity (Supplementary Fig. S1A and Table S2). One of these ORFs was *ASPRV1* that is a known ERV-derived protease gene acquired in the common ancestor of mammals and is responsible for skin maintenance (Matsui et al. 2011). Such ERV-derived ORFs that are annotated as genes in the human genome were removed, and three ORFs remained (Supplementary Table S2). They were located tandemly in the intron of the *GRIP2* gene in the opposite direction (Fig. 1B). All three ORFs showed high similarity to the reverse transcriptase (RT) of spumaretrovirus in GyDB (Supplementary Table S3). Therefore, we designated these gene candidates as *RTOM* [RT-like ORF in Monotreme], and the three were named *RTOM1*, *RTOM2*, and *RTOM3* in order of their location from the 5′ direction (Fig. 1B). To further examine the genomic loci of the *RTOM* ORFs in monotremes, We performed self-alignment of the GRIP2 gene including the three ORFs for the platypus and echidna genomes using LAST program (Kiełbasa et al. 2011). The dot-plots indicate that the three ORFs, including the surrounding regions, were aligned as tandem repeats (Supplementary Fig. S2). We also attempted to align platypus and echidna *GRIP2* with the human and mouse *GRIP2* to gain more insights into the structural evolution of this region including the *RTOM* ORFs; however, the introns of *GRIP2* were not conserved among them, suggesting that the *RTOM* ORFs could emerge in the ancestor of platypus and echidna (Supplementary Fig. S3).

To examine the possibility that the *RTOM* ORFs were acquired before the divergence of therians and monotremes, the nucleotide and amino acid sequences of the *RTOM* ORFs were searched using BLASTn and tBLASTn, respectively, with an e-value < 1E − 5 against all genomes of mammals, birds, and reptiles available in the National Center for Biotechnology Information (NCBI) Assembly. The BLASTn search resulted in significant hits from several

genomes (mammals: 26 out of 510 genomes, birds: 29 out of 556 genomes, and reptiles: eight out of seventy-nine genomes) (Supplementary Table S4); however, their query cover rates were low (up to 5.2 per cent) except for hits to the *RTOM* ORFs themselves. The tBLASTn search identified up to thousands of hits for each genome (Supplementary Table S4). This is because the amino acid sequences of the *RTOM* ORFs are similar to the RT region of other ERVs. We examined the proximity of these hits to the *GRIP2* gene and found no hits considered to be orthologs of the *RTOM* ORFs (see the 'Materials and Methods' section). Therefore, we conclude that the *RTOM* ORFs are monotreme-specific.

We found that, in the platypus genome, there are computationally annotated RefSeq genes containing the *RTOM* ORFs (Fig. 2A). *RTOM1*, *RTOM2*, and *RTOM3* genes of platypus contain two introns in the 5′ UTR, and the entire *RTOM* ORFs are expressed as mRNA excluding a second splicing variant of *RTOM3* that partially lost its ORF (Fig. 2A). In echidna, *RTOM2* and *RTOM3* gene structures were annotated in the RefSeq transcripts; however, *RTOM1* was not annotated. By conducting transcriptome assemblies of RNA-seq data of echidna tissues (Supplementary Table S5), we reconstructed all *RTOM* transcripts including the *RTOM1* (Fig. 2B; Supplementary Fig. S4). We also found a chimeric transcript of *RTOM2* and *RTOM3*, which was transcribed from the transcription start site of *RTOM2*, but its CDS is *RTOM3* (Supplementary Fig. S4). Except for this chimeric transcript, all echidna *RTOM* transcripts have two introns in the 5′ UTR, which was similar to those of platypus. We then constructed a multiple alignment of the seven amino acid sequences of platypus and echidna RTOMs, including two splicing variants of platypus *RTOM3* (Fig. 2C). The amino acid sequence of RTOM2 lacks a region shared by RTOM1 and RTOM3, but the C-terminal region was conserved among the amino acid sequences of RTOMs without insertion or deletion (Fig. 2C). To investigate the tissue-specific expression of the *RTOM* genes, we analyzed the RNA-seq data of platypus and echidna (Supplementary Table S5). In platypus, *RTOM1*, *RTOM2*, and *RTOM3* were commonly highly expressed in the testis (Fig. 2D). *GRIP2* was expressed not only in the testis but also in the brain, and its expression level was lower than that of the *RTOM* genes. We further investigated the mapped reads using Interactive Genome Viewer (Thorvaldsdóttir, Robinson, and Mesirov 2013) (Supplementary Fig. S5) and found that *RTOM3* showed a splicing variant with an intron in the coding region, as shown in the RefSeq transcript. In echidna, we found that all *RTOM* transcripts were specifically expressed in the testis, similar to platypus. Expression of *GRIP2* in echidna testis was also relatively low (Fig. 2E). This suggests that the *RTOM* genes are more actively transcribed than *GRIP2*, or the *RTOM* transcripts are more stable than the *GRIP2* transcripts in the testis. Given the higher expression level of *RTOM2* in both platypus and echidna, this gene may play a central role in the *RTOM* genes. It is still possible that the relative expression levels of three genes may change according to other tissues and developmental stages that were not examined in this study. In addition, since this study did not present the evidence of the translation of the *RTOM* transcripts, whether putative RTOM proteins are involved in testicular function needs to be verified in the future.

To obtain insights into the viral origin of the *RTOM* genes, we performed a BLASTp search of the amino acid sequence of platypus RTOM1 against the NCBI virus database. We found that retrovirus Pol proteins from various distinct lineages, namely gammaretrovirus, deltaretrovirus, epsilonretrovirus, and spumaretrovirus, are similar to the amino acid sequence of RTOM1 (BLASTp: E-value < 1E − 20). In all hits, the retroviral Pol proteins showed high similarity to the latter half of RTOM1 (approximately 370-607aa) (Fig. 3A). A domain search against the Pfam database (Mistry et al. 2021) in the HMMER web service (Finn, Clements, and Eddy 2011) revealed that the latter half of RTOM1 and RTOM3 contain RT domains (Supplementary Fig. S6). A phylogenetic tree was constructed from the RT regions of the amino acid sequences of RTOMs and the retroviral Pol proteins (Fig. 3B). The amino acid sequences of RTOMs appear to be more related to class III retroviruses, including spumaviruses or spumavirus-related MuERV-L (Llorens et al. 2009). The tree topology of the RTOMs strongly suggests that all the *RTOM* genes were formed before the divergence of platypus and echidna. Together with the self-alignment of genomic sequences (Supplementary Fig. S2), three *RTOM* genes were generated by tandem gene duplications before the divergence of platypus and echidna (Fig. 3C). In the non-RT region of the amino acid sequences of RTOM1 (approximately 1–369aa), no significant hits for retroviruses were obtained (Fig. 3A). We performed a BLASTp search for all non-redundant proteins in the GenBank database for the non-RT region of the amino acid sequences of RTOM1; however, no similar proteins were found except for the putative proteins of RTOM2 and RTOM3 (E-value < 0.05). This suggests that the non-RT region was derived from a non-retroviral sequence or was derived from the retroviral gene that has accumulated too many mutations to be aligned with retroviruses. Considering the structural divergence of the non-RT region, such as deletion of RTOM2 and splicing variant of platypus RTOM3 (Fig. 2C), the RT region is a core domain of the putative RTOM proteins, and the non-RT region may provide functional modifications specific to each putative RTOM protein.

During the 187-million-year history of diverging from monotremes, therians have acquired many ERV genes and evolved their unique features, especially the placenta (Imakawa and Nakagawa 2017). Our work revealed that monotremes also domesticated ERV genes that emerged and were conserved more than 55 million years ago, the divergence time of platypus and echidna (Zhou et al. 2021). Although the translation of *RTOM* genes was not confirmed in this study, the calculation of nonsynonymous and synonymous nucleotide substitution frequencies of *RTOM1*, *RTOM2*, and *RTOM3* shows that their amino acid sequences are under purifying selection, strongly suggesting that they physiologically function as proteins (Fig. 3D). We found that the RT domain of all putative RTOM proteins lacked the three catalytic carboxylates of aspartic acids (Supplementary Fig. S7). These amino acid residues are highly conserved among all retroviruses, and the replacement of these amino acids results in a complete loss of the RT activity (Larder et al. 1987; Sarafianos et al. 2009). Therefore, the putative RTOM proteins may have different functions from those of reverse transcription. To the best of our knowledge, there are no retroviral genes in which only the RT domain is co-opted in vertebrates (Naville et al. 2016). Although this study has a limitation that the expression and function of the putative RTOM proteins have not been fully validated due to difficulties in obtaining tissues, the future functional elucidation of *RTOM1*, *RTOM2*, and *RTOM3* will provide us with new aspects of ERV-derived genes functioning in mammals.

Materials and methods

Identification of conserved ERV genes

The platypus genome (mOrnAna1.p.v1, GCF_004115215.1) and the echidna genome (mTacAcu1.pri, GCF_015852505.1) were used for the ERV gene screening. The 240-nt ORF flanked by stop codons

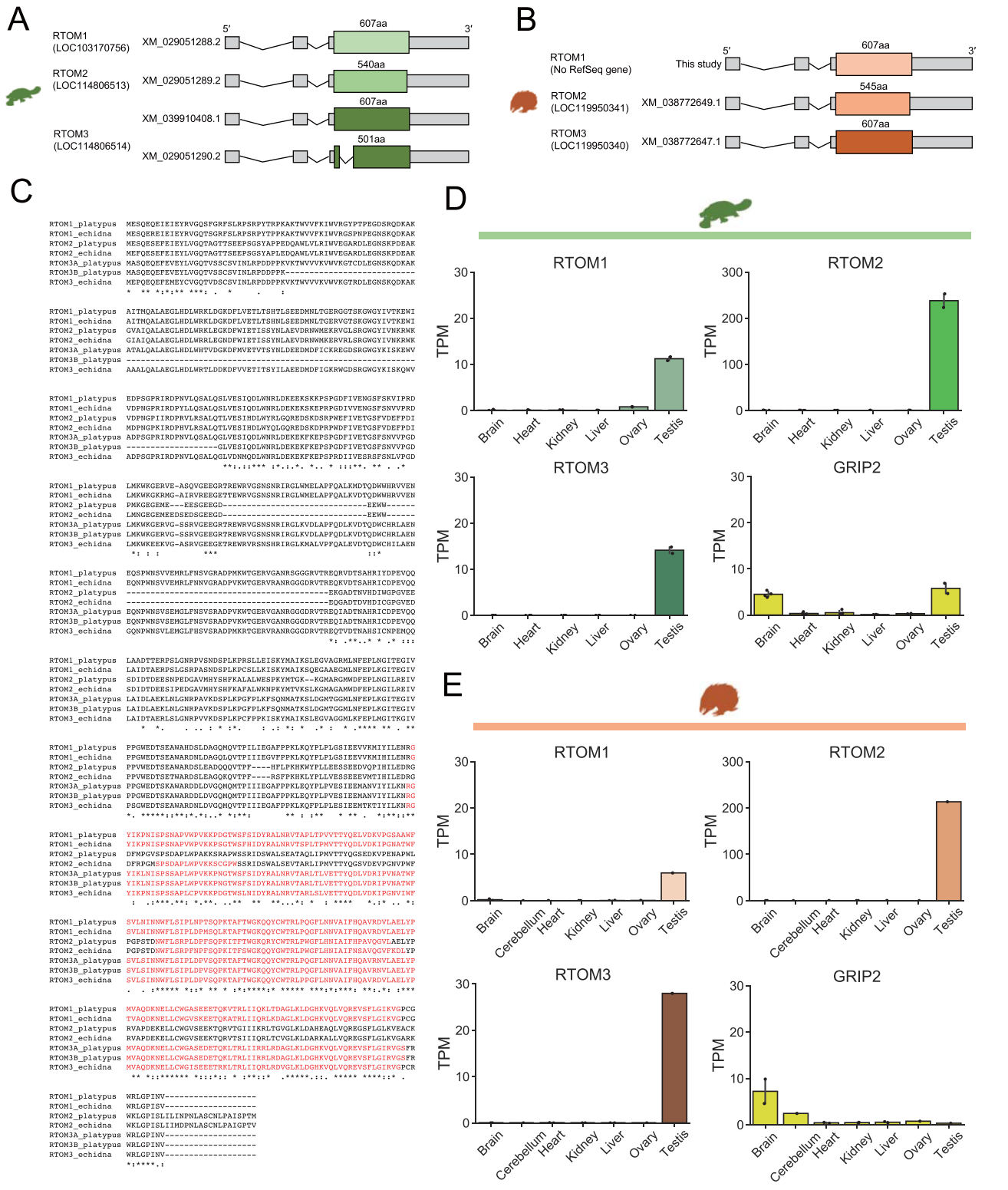
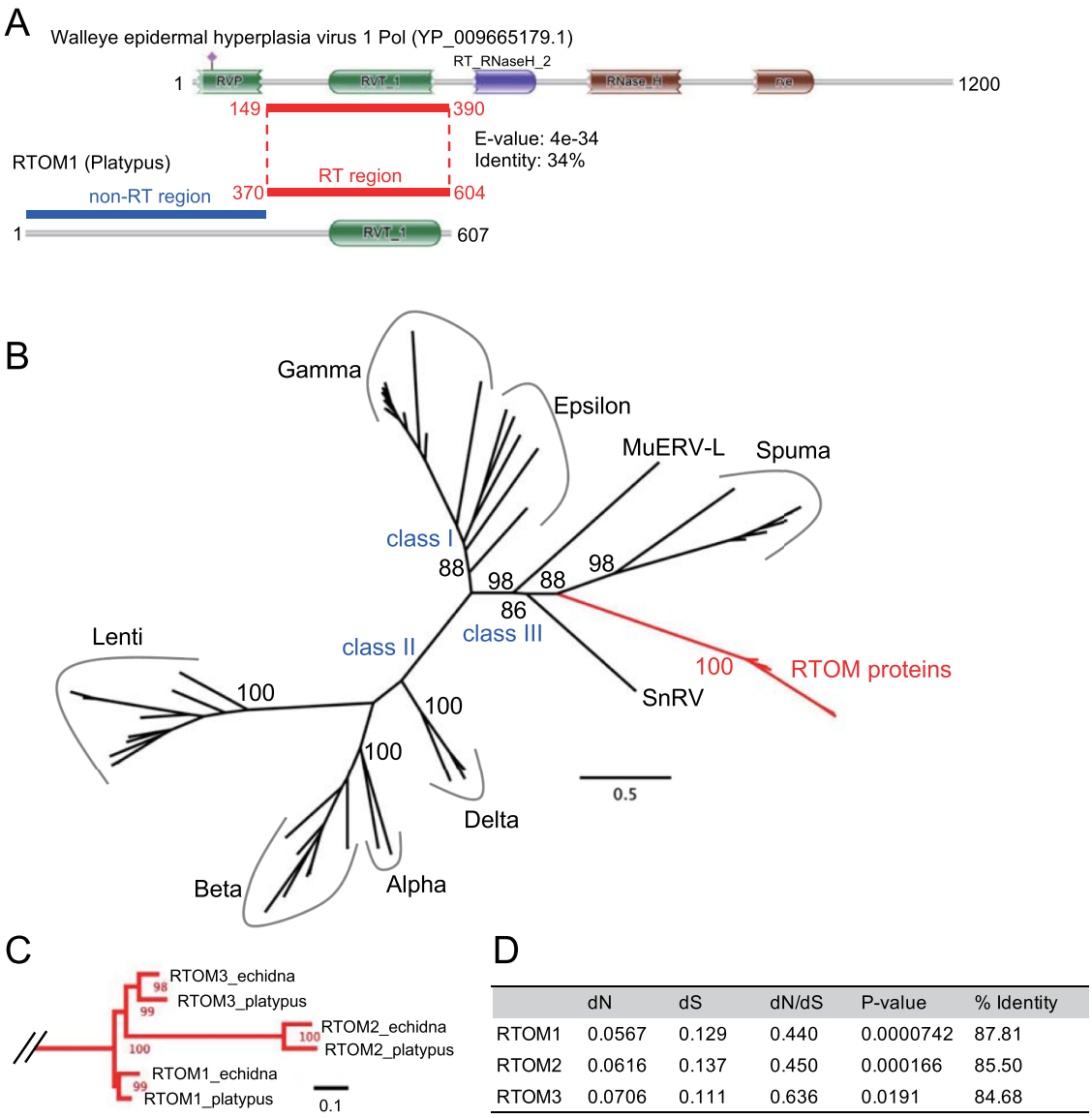


Figure 2. Expression of *RTOM1*, *RTOM2*, and *RTOM3*. (A) Schematic representation of the RefSeq transcripts of the *RTOM* genes in platypus. (B) Schematic representation of the reconstructed *RTOM1* transcript and RefSeq transcripts of the *RTOM2* and *RTOM3* genes in echidna. (C) Multiple alignment of the amino acid sequences of RTOMs. The amino acid sequence of echidna RTOM1 was obtained from the genomic ORF. 'RTOM3A_plasypus' and 'RTOM3B_platypus' are isoforms derived from 'XM_039910408.1' and 'XM_029051290.2,' respectively. The regions showing similarity to the HMM of spumaretrovirus RT domain in GyDB are indicated in red. (D, E) Tissue-specific expression of *RTOM* genes and *GRIP2* in (D) platypus and (E) echidna. Normalized expression levels are presented as transcript per million (TPM).

was retrieved using the getorf program in the European Molecular Biology Open Software Suite (Rice, Longden, and Bleasby 2000). For HMM-based sequence search, hmmscan was used (expected threshold: 1E – 5) in HMMER3 v3.2.2 (Eddy 2011). ORFs were clustered using CD-HIT v4.8.1 (Li and Godzik 2006) with 50 per cent amino acid identity. The sequence search for platypus



	dN	dS	dN/dS	P-value	% Identity
RTOM1	0.0567	0.129	0.440	0.0000742	87.81
RTOM2	0.0616	0.137	0.450	0.000166	85.50
RTOM3	0.0706	0.111	0.636	0.0191	84.68

Figure 3. Evolution of *RTOM1*, *RTOM2*, and *RTOM3*. (A) Comparison between platypus RTOM1 and retroviral Pol protein. Walleye epidermal hyperplasia virus 1 is represented as an example. A region showing similarity to the Pol protein by BLASTp was designated as 'RT region'. A region that did not show similarity to any retroviral genes was designated as 'non-RT region'. (B) A phylogenetic tree constructed from the amino acid sequences of RT regions of the six RTOMs and the retroviral Pol proteins in GyDB. The multiple alignment is available in Supplementary File 3. Ultrafast-bootstrap values obtained from 1000 times replication are shown in major branches. (C) Detailed representation of the clade of the amino acid sequences of RTOMs. The scale was shown in the right bottom. (D) The numbers of nonsynonymous substitutions (dN) and synonymous substitutions (dS) per site estimated by Nei–Gojobori method (Nei and Gojobori 1986). Statistical significance of selection was estimated by the codon-based Z test of neutrality using MEGA-X (Kumar et al. 2018). The % identity was calculated using the Ident and Sim program (https://www.bioinformatics.org/sms2/ident_sim.html) (Stothard 2000).

ORFs against echidna ORFs was conducted using BLASTp v2.10.0+ with an e-value < 1E – 50 (Camacho et al. 2009). By examining the distribution of the bitscore of the BLASTp search, we extracted ORF pairs that showed high similarity between the two species (Supplementary Fig. S1A). Then, we examined the RefSeq annotation and identified the genes to which the ORFs belonged.

To examine the presence of homologous sequences beyond the monotreme lineage, the genes that were not described in the human RefSeq genes were subjected to deep homology searches. We performed BLASTn and tBLASTn v2.10.0+ with e-values < 1E – 5 against all genomes of mammals, birds, and reptiles downloaded on 22 January 2022 (Supplementary Table S4). The query cover rate for each hit was calculated as [alignment length/query length]. To examine whether the amino acid sequences obtained by tBLASTn are the ortholog of the *RTOM* genes, we investigated the proximity of these hits to the *GRIP2* gene as follows. First, we extracted hits to the amino acid sequences of RTOMs with a query cover of at least 60 per cent. Second, to obtain the genomic position of *GRIP2* on each genome, we performed BLAT v0.35 (Kent 2002) using the amino acid sequence of the human GRIP2 (NP_001073892.3) against the 1,145 genomes. We considered the hit with the highest score in each genome as the *GRIP2* position. Finally, we compared the genomic position of the hits of *RTOM* and *GRIP2* and confirmed that all of them were located on different contigs or were located far enough from each other (the closest pair of hits to *RTOM* and *GRIP2* in the same contig is 7.7 Mb apart from each other).

To validate this approach, we performed similar analyses on the genomes of humans (GRCh38.p13) and marmosets (Callithrix_jacchus_cj1700_1.1) (Supplementary Fig. S1B and C). They diverged 43 million years ago (Perelman et al. 2011). We successfully identified known ERV-derived genes such as *PEG10* (Ono et al. 2001), *RTL1/PEG11* (Charlier et al. 2001), *ASPRV1* (Matsui et al. 2011), *NYNLIN/CGIN1* (Marco and Marín 2009), *ERVV-1* and *2* (Kjeldbjerg et al. 2008), and *ERVMER34-1/HEMO* (Heidmann et al. 2017) (Supplementary Table S6). This suggests that our method is sensitive enough to identify ERV ORFs conserved in platypus and echidna.

Expression analyses

RNA-seq data of platypus (twenty samples from six tissues) (Marin et al. 2017) and echidna (eleven samples from seven tissues) (Zhou et al. 2021) were used (Supplementary Table S5). Low-quality reads were trimmed and filtered using fastp v0.19.5 with default options (Chen et al. 2018). The filtered reads were mapped to each reference genome using HISAT2 v2.1.0 with default option (allowing ≤5 multiple mappings) (Pertea et al. 2016). Based on the eleven RNA-seq sequencing data mapped on the echidna genome, we obtained the echidna *RTOM1* transcript by conducting transcriptome assembly using Stringtie2 v2.1.6 with '–merge' option (Kovaka et al. 2019). We have added the coordinates of the echidna *RTOM1* transcript (Supplementary File 1) to the RefSeq gene coordinates. We then calculated the expression levels for twenty platypus and eleven echidna RNA-seq samples using the Stringtie2 program with default options (Kovaka et al. 2019). To extract unique-mapped reads from a given aligner-generated SAM file, we collected reads containing the 'NH:i:1' flag indicating that they were uniquely mapped into the genome.

Phylogenetic analyses

Representative retroviral Pol amino acid sequences were retrieved from the GyDB collection (https://gydb.org/index.php/Alignment?alignment=POL_retroviridae_Biology_Direct_4_41_2009&format=txt) (Llorens et al. 2009). A multiple alignment was generated using MAFFT v7.487 (Katoh and Standley 2013), and poorly aligned regions were removed using trimAl v1.4.rev15 (Capella-Gutiérrez, Silla-Martínez, and Gabaldón 2009). A phylogenetic tree was constructed using IQ-TREE2 v2.0.8 (Minh et al. 2020) with 1,000 replicates of ultrafast-bootstrap (Hoang et al. 2018). The tree was visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Supplementary data

Supplementary data are available at *Virus Evolution Journal* online.

Acknowledgements

We would like to thank Editage for the English language editing. This work was supported by Grant-in-Aid for JSPS fellows 20J22607 to K.K. and JSPS KAKENHI 20K06775 and 20H03150 to T.M. and S.N. The super-computing resource was partially supported by the National Institute of Genetics (NIG) supercomputer at ROIS NIG.

Conflict of interest: None declared.

References

Best, S. et al. (1996) 'Positional Cloning of the Mouse Retrovirus Restriction Gene Fvl', *Nature*, 382: 826–9.

Blaise, S. et al. (2003) 'Genomewide Screening for Fusogenic Human Endogenous Retrovirus Envelopes Identifies Syncytin 2, a Gene Conserved on Primate Evolution', *Proceedings of the National Academy of Sciences of the United States of America*, 100: 13013–8.

Blond, J.-L. et al. (2000) 'An Envelope Glycoprotein of the Human Endogenous Retrovirus HERV-W Is Expressed in the Human Placenta and Fuses Cells Expressing the Type D Mammalian Retrovirus Receptor', *Journal of Virology*, 74: 3321–9.

Boso, G. et al. (2021) 'The Oldest Co-opted Gag Gene of a Human Endogenous Retrovirus Shows Placenta-Specific Expression and Is Upregulated in Diffuse Large B-Cell Lymphomas', *Molecular Biology and Evolution*, 38: 5453–71.

Camacho, C. et al. (2009) 'BLAST+: Architecture and Applications', *BMC Bioinformatics*, 10: 421.

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009) 'trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses', *Bioinformatics*, 25: 1972–3.

Charlier, C. et al. (2001) 'Human–Ovine Comparative Sequencing of a 250-kb Imprinted Domain Encompassing the Callipyge (Clpg) Locus and Identification of Six Imprinted Transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8', *Genome Research*, 11: 850–62.

Chen, S. et al. (2018) 'fastp: An Ultra-fast All-in-One FASTQ Preprocessor', *Bioinformatics*, 34: i884–90.

Dupressoir, A. et al. (2009) 'Syncytin-A Knockout Mice Demonstrate the Critical Role in Placentation of a Fusogenic, Endogenous Retrovirus-Derived, Envelope Gene', *Proceedings of the National Academy of Sciences of the United States of America*, 106: 12127–32.

Dupressoir, A. et al. (2011) 'A Pair of Co-opted Retroviral Envelope Syncytin Genes Is Required for Formation of the Two-layered Murine Placental Syncytiotrophoblast', *Proceedings of the National Academy of Sciences*, 108: E1164–73.

Eddy, S. R. (2011) 'Accelerated Profile HMM Searches', *PLoS Computational Biology*, 7: e1002195.

Emerson, R. O., and Thomas, J. H. (2011) 'Gypsy and the Birth of the SCAN Domain', *Journal of Virology*, 85: 12043–52.

Finn, R. D., Clements, J., and Eddy, S. R. (2011) 'HMMER Web Server: Interactive Sequence Similarity Searching,' *Nucleic Acids Research*, 39(suppl): W29–37.

Heidmann, O. et al. (2017) 'HEMO, an Ancestral Endogenous Retroviral Envelope Protein Shed in the Blood of Pregnant Women and Expressed in Pluripotent Stem Cells and Tumors', *Proceedings of the National Academy of Sciences*, 114: E6642–51.

Hoang, D. T. et al. (2018) 'UFBoot2: Improving the Ultrafast Bootstrap Approximation', *Molecular Biology and Evolution*, 35: 518–22.

Ikeda, H., and Sugimura, H. (1989) 'Fv-4 Resistance Gene: A Truncated Endogenous Murine Leukemia Virus with Ecotropic Interference Properties', *Journal of Virology*, 63: 5405–12.

Imakawa, K., and Nakagawa, S. (2017) 'The Phylogeny of Placental Evolution through Dynamic Integrations of Retrotransposons', *Progress in Molecular Biology and Translational Science*, 145: 89–109.

Imakawa, K., Nakagawa, S., and Miyazawa, T. (2015) 'Baton Pass Hypothesis: Successive Incorporation of Unconserved Endogenous Retroviral Genes for Placentation during Mammalian Evolution', *Genes to Cells*, 20: 771–88.

Johnson, W. E. (2019) 'Origins and Evolutionary Consequences of Ancient Endogenous Retroviruses', *Nature Reviews. Microbiology*, 17: 355–70.

Katoh, K., and Standley, D. M. (2013) 'MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability', *Molecular Biology and Evolution*, 30: 772–80.

Kent, W. J. (2002) 'BLAT—The BLAST-Like Alignment Tool', *Genome Research*, 12: 656–64.

Kiełbasa, S. M. et al. (2011) 'Adaptive Seeds Tame Genomic Sequence Comparison', *Genome Research*, 21: 487–93.

Kjeldbjerg, A. L. et al. (2008) 'Gene Conversion and Purifying Selection of a Placenta-Specific ERV-V Envelope Gene during Simian Evolution', *BMC Evolutionary Biology*, 8: 266.

Kovaka, S. et al. (2019) 'Transcriptome Assembly from Long-read RNA-seq Alignments with StringTie2', *Genome Biology*, 20: 278.

Kumar, S. et al. (2018) 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms', *Molecular Biology and Evolution*, 35: 1547–9.

Larder, B. A. et al. (1987) 'Site-Specific Mutagenesis of AIDS Virus Reverse Transcriptase', *Nature*, 327: 716–7.

Li, W., and Godzik, A. (2006) 'Cd-hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences', *Bioinformatics*, 22: 1658–9.

Llorens, C. et al. (2011) 'The Gypsy Database (Gydb) of Mobile Genetic Elements: Release 2.0', *Nucleic Acids Research*, 39(Database): D70–74.

Llorens, C. et al. (2009) 'Network Dynamics of Eukaryotic LTR Retroelements beyond Phylogenetic Trees', *Biology Direct*, 4: 41.

Marco, A., and Marin, I. (2009) 'CGIN1: A Retroviral Contribution to Mammalian Genomes', *Molecular Biology and Evolution*, 26: 2167–70.

Marin, R. et al. (2017) 'Convergent Origination of a Drosophila-like Dosage Compensation Mechanism in a Reptile Lineage', *Genome Research*, 27: 1974–87.

Matsui, T. et al. (2011) 'SASPase Regulates Stratum Corneum Hydration through Profilaggrin-to-Filaggrin Processing', *EMBO Molecular Medicine*, 3: 320–33.

Mi, S. et al. (2000) 'Syncytin Is a Captive Retroviral Envelope Protein Involved in Human Placental Morphogenesis', *Nature*, 403: 785–9.

Minh, B. Q. et al. (2020) 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Molecular Biology and Evolution*, 37: 1530–4.

Mistry, J. et al. (2021) 'Pfam: The Protein Families Database in 2021', *Nucleic Acids Research*, 49: D412–9.

Nakagawa, S., and Takahashi, M. U. (2016) 'gEVE: A Genome-based Endogenous Viral Element Database Provides Comprehensive Viral Protein-Coding Sequences in Mammalian Genomes', *Database*, 2016: 1–8.

Naville, M. et al. (2016) 'Not so Bad after All: Retroviruses and Long Terminal Repeat Retrotransposons as a Source of New Genes in Vertebrates', *Clinical Microbiology and Infection*, 22: 312–23.

Nei, M., and Gojobori, T. (1986) 'Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions', *Molecular Biology and Evolution*, 3: 418–26.

Ono, R. et al. (2001) 'A Retrotransposon-Derived Gene, PEG10, Is a Novel Imprinted Gene Located on Human Chromosome 7aq21', *Genomics*, 73: 232–7.

Perelman, P. et al. (2011) 'A Molecular Phylogeny of Living Primates', *PLoS Genetics*, 7: 1–17.

Pertea, M. et al. (2016) 'Transcript-Level Expression Analysis of RNA-seq Experiments with HISAT, StringTie and Ballgown', *Nature Protocols*, 11: 1650–67.

Rice, P., Longden, L., and Bleasby, A. (2000) 'EMBOSS: The European Molecular Biology Open Software Suite', *Trends in Genetics*, 16: 276–7.

Sarafianos, S. G. et al. (2009) 'Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition', *Journal of Molecular Biology*, 385: 693–713.

Stothard, P. (2000) 'The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences', *BioTechniques*, 28: 1102–4.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013) 'Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration', *Briefings in Bioinformatics*, 14: 178–92.

Ueda, M. T. et al. (2020) 'Comprehensive Genomic Analysis Reveals Dynamic Evolution of Endogenous Retroviruses that Code for Retroviral-like Protein Domains', *Mobile DNA*, 11: 29.

Wang, J., and Han, G. Z. (2020) 'Frequent Retroviral Gene Co-option during the Evolution of Vertebrates', *Molecular Biology and Evolution*, 37: 3232–42.

Warren, W. C. et al. (2008) 'Genome Analysis of the Platypus Reveals Unique Signatures of Evolution', *Nature*, 453: 175–83.

Zhou, Y. et al. (2021) 'Platypus and Echidna Genomes Reveal Mammalian Biology and Evolution', *Nature*, 592: 756–62.