Letter

Discoveries

Title

Monotreme-specific conserved proteins derived from retroviral reverse transcriptase

Authors

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

Affiliations

[1]Laboratory of Virus-Host Coevolution, Institute for Frontier Life and Medical Sciences, Kyoto University, Sakyo-ku, Kyoto 606-8507, Japan

[2]Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Kanagawa 259-1193, Japan.

*Correspondence to: takavet@infront.kyoto-u.ac.jp (T.M.) and so@tokai.ac.jp (S.N.)

Abstract

Endogenous retroviruses (ERVs) have played an essential role in the evolution of mammals. Many ERV-derived genes are reported in the therians that 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*, *2*, and *3*. They were found to be clustered in the *GRIP2* intron. Phylogenetic analysis revealed that *RTOM1*, *2*, and *3* are strongly conserved between these species, and they were generated by tandem duplications before the divergence of platypus and echidna. The *RTOM* transcripts were specifically expressed in the testis, suggesting the physiological importance of *RTOM* genes. 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.

32 Endogenous retroviruses (ERVs) are remnants of retroviral genomes found in the host
33 genomes. ERVs are retroviruses that infect the host germline cells and are integrated into
34 the host genome (Johnson 2019). Young ERVs retain their viral open reading frames
35 (ORFs), but gradually lose their intact ORFs due to the accumulation of mutations.
36 However, proteins expressed from ERVs sometimes evolve as functional genes in the host
37 (Ueda et al. 2020). A typical example is the syncytin genes that are derived from ERVs.
38 In humans, syncytin-1 and -2 encode fusogenic proteins that are expressed in the placenta
39 (Mi et al. 2000; Blond et al. 2000; Blaise et al. 2003). In mice, other ERV-derived
40 fusogenic genes, syncytin-A and -B, are required for trophoblast cell fusion in the placenta
41 (Dupressoir et al. 2009; Dupressoir et al. 2011). In mammals, such various fusogenic
42 genes are known to be involved in placental development, which have been independently
43 acquired from different ERVs in different mammalian lineages. That is a representative
44 example of the convergent evolution (Imakawa et al. 2015). In addition, other ERV-
45 derived genes that does not show fusogenic activity have also been found to be expressed
46 in the placenta. For example, *HEMO* encoding a secreted envelope protein (Heidmann et
47 al. 2017) as well as *gagV1* and *pre-gagV1* genes (Boso et al. 2021) are highly expressed
48 in the human placenta. In contrast, it is unknown whether ERV-derived genes are co-opted
49 in monotremes that are egg-laying mammals. Comparative studies for the detection of
50 ERV-derived genes have been conducted in mammalian genomes, including the platypus
51 (Nakagawa and Takahashi 2016; Wang and Han 2020). However, for monotremes, only
52 the genome sequence of one species, the platypus, was available (OANA5), the quality
53 of which was limited (Warren et al. 2008). Recently, high-quality monotreme genomes of
54 platypus (mOrnAna1.p.v1) and echidna (mTacAcu1.pri) were sequenced using long-read
55 sequencing technology (Zhou et al. 2021), making it possible to search for conserved

56 ERV genes in monotremes. Here, we performed a comparative analysis of the genomes
57 of these two monotremes and found three ERV-derived genes specific to the monotreme
58 lineage.
59
60 To comprehensively search for ERV genes in monotremes, we extracted ORFs from the
61 genomes of platypus and echidna. The amino acid sequences obtained by the virtual
62 translation of these ORFs were used as queries for the sequence search. We used the
63 hidden Markov model (HMM) of the retroviral genes in the Gypsy Database 2.0 (GyDB)
64 (Llorens et al. 2011) as the subject of the sequence search (supplementary table S1). We
65 identified ORFs similar to *gag*, *pro*, *pol*, and *env* genes (fig. 1A). These ORFs are
66 presumed to be a mixture of (1) ORFs that are evolutionarily conserved and (2) ORFs of
67 young transposons that retain their ORFs. To exclude young ERV ORFs, we performed
68 the clustering analysis based on the amino acid sequence identity. Since young ERVs are
69 thought to be included in large clusters due to their mutual similarity, we removed
70 sequences that belonged to large clusters consisting of more than 10 sequences. Next,
71 using the platypus ORFs as queries, and the echidna ORFs as the subjects, we conducted
72 a sequence similarity search using BLASTp. We obtained nine ORF pairs with high amino
73 acid similarity and the same synteny between platypus and echidna (supplementary table
74 S2). For six out of the nine ORF pairs, we found respective homologs in the human
75 genome, indicating that they were either ERV genes acquired in the common mammalian
76 ancestor or host genes with high similarity to ERVs. Indeed, one of the six genes is
77 *ASPRV1* that is a known ERV-derived protease gene acquired in the common ancestor of
78 mammals and is responsible for skin maintenance (Matsui et al. 2011). The remaining
79 three genes were not found in the human genome. They were located tandemly in the

80 intron of the *GRIP2* gene in the opposite direction (fig. 1B). All three ORFs showed high
81 similarity to the reverse transcriptase (RT) of spumaretrovirus in GyDB (supplementary
82 table S3). Therefore, we designated these genes as *RTOM* [RT-like ORF in Monotreme],
83 and three genes were named as *RTOM1*, *RTOM2*, and *RTOM3* in order of their location
84 from the 5′ direction (fig. 1B). The *RTOM* coding sequences were searched in the
85 genomes of six mammals, two birds, eight reptilians, and two amphibians (supplementary
86 table S4); however, significant hits were not obtained other than in platypus and echidna
87 (BLASTn: E-value < 1E-5). Therefore, the *RTOM* genes could be monotreme-specific
88 and originated more than 55 million years ago, the divergence time of platypus and
89 echidna (Zhou et al. 2021).
90
91 We found that the gene structures of *RTOM* genes in the platypus genome were annotated
92 in the RefSeq database (fig. 2A). *RTOM1*, *2*, and *3* genes of platypus contained two
93 introns in the 5′ UTR, and the entire *RTOM* ORFs are expressed as mRNA excluding a
94 second splicing variant of *RTOM3* that partially lost its ORF (fig. 2A). In echidna,
95 *RTOM2* and *RTOM3* gene structures were annotated in the RefSeq transcripts; however,
96 *RTOM1* was not annotated. By conducting transcriptome assemblies of RNA-seq data of
97 echidna tissues (supplementary table S5), we reconstructed the *RTOM1* transcript (fig.
98 2B; supplementary data S1). As a result, all echidna *RTOM* transcripts have two introns
99 in the 5′ UTR, which was similar to observations for platypus. According to the alignment
100 of the six amino acid sequences of platypus and echidna RTOM proteins, RTOM2 lacks
101 a region shared by RTOM1 and RTOM3, but the C-terminal region was conserved among
102 the RTOM proteins without insertion or deletion (fig. 2C). To investigate the tissue-
103 specific expression of *RTOM* genes, we analyzed the RNA-seq data of platypus and

echidna (supplementary table S5). In platypus, *RTOM1*, *2*, and *3* 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. This suggests that the *RTOM* expression was not a result of the *GRIP2* expression. We further investigated the mapped reads using Interactive Genome Viewer (Thorvaldsdóttir et al. 2013). It was found that *RTOM3* showed a splicing variant with an intron in the coding region, as shown in the RefSeq transcript (supplementary fig S1). 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, strengthening the idea that the *RTOM* expression is independent of *GRIP2* expression (fig. 2E). Given the higher expression level of *RTOM2* in both platypus and echidna, this gene may play a central role of the RTOM proteins. It is still possible that the relative expression levels of three genes may change according to tissues and developmental stages that were not examined in this study.

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 RTOM1 proteins (BLASTp: E-value < 1E-20). In all hits, the retroviral Pol proteins showed high similarity to the latter half of RTOM1 (about 370-607aa). Domain search against the Pfam database (Mistry et al. 2021) in the HMMER web service (Finn et al. 2011) revealed that the latter half of RTOM1 and RTOM3 contain RT domains (fig. 3A; supplementary fig. S2). A phylogenetic tree was constructed from the RT regions of the

128 RTOM proteins and the retroviral Pol proteins (fig. 3B). The RTOM proteins appear to
129 be more related to class III retroviruses, including spumaviruses or spumavirus-related
130 MuERV-L (Llorens et al. 2009). The tree topology suggested that *RTOM1* emerged at first,
131 and *RTOM2* and *RTOM3* were then generated by tandem gene duplications before the
132 divergence of platypus and echidna (fig. 3C). In the non-RT region of RTOM1
133 (approximately 1-369aa), no significant hits for retroviruses were obtained (fig. 3A). We
134 performed a BLASTp search for all non-redundant proteins in the GenBank database for
135 the non-RT region of RTOM1; however, no similar proteins were found except for
136 RTOM2 and 3 (E-value < 0.05). Therefore, the non-RT region of the *RTOM* genes does
137 not seem to be derived from ERV genes or conserved host genes. Considering the
138 structural divergence of the non-RT region, such as deletion of RTOM2 and splicing
139 variant of platypus RTOM3 (fig. 2C), the RT region is a core domain of the RTOM
140 proteins, and the non-RT region may provide functional modifications specific to each
141 RTOM protein.
142
143 During the 187-million-years history after diverging from monotremes (Zhou et al. 2021),
144 therians have acquired many ERV genes and evolved their unique features, especially the
145 placenta (Imakawa and Nakagawa 2017). Our work revealed that monotremes also
146 domesticated ERV genes. The functional inference of the RTOM proteins is difficult as
147 co-opted RT genes genetically similar to the *RTOM* genes have not been reported in other
148 vertebrates to the best of our knowledge. One possibility is that RTOM proteins may
149 function as restriction factors against ERVs and retrotransposons as well as exogenous
150 retroviruses. For example, gag-derived *Fv1* (Best et al. 1996) and env-derived *Fv4* (Ikeda
151 and Sugimura 1989) inhibit retroviral infection in mice. It is possible that the RT domains

152 in the *RTOM* genes may compete with retroviral infection and/or retrotransposition as
153 antagonists. Another possibility is that RTOM proteins are involved in physiological
154 functions unique to monotremes. To investigate that, it would be important to clarify
155 which cells (viz. germ cells or somatic cells) in testis express the *RTOM* genes. Further
156 studies pertaining to *RTOM1*, *2*, and *3* in platypus and echidna will expand our
157 understanding of ERV co-option during the evolution of mammals.
158

159 **Materials and Methods**

160 *Identification of conserved ERV genes*

161 The platypus genome (mOrnAna1.p.v1, GCF_004115215.1) and the echidna genome
162 (mTacAcu1.pri, GCF_015852505.1) were used for the ERV gene screening (please see
163 fig. 1). The 240-nt ORF flanked by stop codons were retrieved using the getorf program
164 in the European Molecular Biology Open Software Suite (Rice et al. 2000). For HMM-
165 based sequence search, hmmscan was used (Expected threshold: 1E-5) in HMMER3
166 v3.2.2 (Eddy 2011). ORFs were clustered using CD-HIT v4.8.1 (Li and Godzik 2006)
167 with 50% amino acid identity. The sequence search for platypus ORFs against echidna
168 ORFs was conducted using BLASTp v2.10.0+ with an e-value < 1E-50 (Camacho et al.
169 2009). Hits with a bitscore > 400 were retrieved and checked the synteny was checked
170 using the UCSC genome browser (https://genome.ucsc.edu/index.html).
171

172 *Expression analysis*

173 RNA-seq data of platypus (20 samples from 6 tissues) (Marin et al. 2017) and echidna
174 (11 samples from 7 tissues) (Zhou et al. 2021) were used (supplementary file S5). Low-
175 quality reads were trimmed and filtered using fastp v0.19.5 with default options (Chen et

al. 2018). The filtered reads were mapped to the each reference genome using HISAT2 v2.1.0 (Pertea et al. 2016). Based on the 11 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 added the coordinates of the echidna *RTOM1* transcript (supplementary data S1) to the RefSeq gene coordinates. We then calculated the expression levels for 20 platypus and 11 echidna RNA-seq samples using the Stringtie2 program with default options (Kovaka et al. 2019).

Phylogenetic analysis

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 et al. 2009). A phylogenetic tree was constructed using IQ-TREE2 v2.0.8 (Minh et al. 2020) with 1000 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/).

Acknowledgments

We would like to thank Editage for 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 NIG supercomputer at ROIS National Institute of Genetics.

References

Best S, Le Tissier P, Towers G, Stoye JP. 1996. Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature*. 382:826–829.

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

Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, Mandrand B, Mallet F, Cosset FL. 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. *J Virol.* 74:3321–3329.

Boso G, Fleck K, Carley S, Liu Q, Buckler-White A, Kozak CA. 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. *Mol Biol Evol.* 38:5453–5471.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890.

224 Dupressoir A, Vernochet C, Bawa O, Harper F, Pierron G, Opolon P, Heidmann T.
225 2009. Syncytin-A knockout mice demonstrate the critical role in placentation of a
226 fusogenic, endogenous retrovirus-derived, envelope gene. *Proc Natl Acad Sci U*
227 *S A.* 106:12127–12132.

228 Dupressoir A, Vernochet C, Harper F, Guégan J, Dessen P, Pierron G, Heidmann T.
229 2011. A pair of co-opted retroviral envelope syncytin genes is required for
230 formation of the two-layered murine placental syncytiotrophoblast. *Proc Natl*
231 *Acad Sci U S A* 108:E1164–E1173.

232 Eddy SR. 2011. Accelerated profile HMM Searches. *PLoS Comput Biol.* 7:e1002195.

233 Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence
234 similarity searching. *Nucleic Acids Res.* 39:W29–W37.

235 Heidmann O, Béguin A, Paternina J, Berthier R, Deloger M, Bawa O, Heidmann T.
236 2017. HEMO, an ancestral endogenous retroviral envelope protein shed in the
237 blood of pregnant women and expressed in pluripotent stem cells and tumors.
238 *Proc Natl Acad Sci U S A.* 114:E6642–E6651.

239 Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2:
240 Improving the ultrafast bootstrap approximation. *Mol Biol Evol.* 35:518–522.

241 Ikeda H, Sugimura H. 1989. Fv-4 resistance gene: a truncated endogenous murine
242 leukemia virus with ecotropic interference properties. *J. Virol.* 63:5405–5412.

243 Imakawa K, Nakagawa S. 2017. The phylogeny of placental evolution through dynamic
244 integrations of retrotransposons. *Prog Mol Biol Transl Sci.* 145:89–109.

245 Imakawa K, Nakagawa S, Miyazawa T. 2015. Baton pass hypothesis: successive
246 incorporation of unconserved endogenous retroviral genes for placentation during
247 mammalian evolution. *Genes Cells.* 20:771–788.

248 Johnson WE. 2019. Origins and evolutionary consequences of ancient endogenous
249 retroviruses. *Nat Rev Microbiol.* 17:355–370.

250 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
251 improvements in performance and usability. *Mol Biol Evol.* 30:772–780.

252 Kovaka S, Zimin A V., Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019.

253 Transcriptome assembly from long-read RNA-seq alignments with StringTie2.
254 *Genome Biol.* 20:278.

255 Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of
256 protein or nucleotide sequences. *Bioinformatics.* 22:1658–1659.

257 Llorens C, Muñoz-Pomer A, Bernad L, Botella H, Moya A. 2009. Network dynamics of
258 eukaryotic LTR retroelements beyond phylogenetic trees. *Biol Direct* 4:41.

259 Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-
260 Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP, et al. 2011. The Gypsy
261 Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.*
262 39:D70–D74.

263 Marin R, Cortez D, Lamanna F, Pradeepa MM, Leushkin E, Julien P, Liechti A, Halbert
264 J, Brüning T, Mössinger K, et al. 2017. Convergent origination of a Drosophila-
265 like dosage compensation mechanism in a reptile lineage. *Genome Res.* 27:1974–
266 1987.

267 Matsui T, Miyamoto K, Kubo A, Kawasaki H, Ebihara T, Hata K, Tanahashi S,
268 Ichinose S, Imoto I, Inazawa J, et al. 2011. SASPase regulates stratum corneum
269 hydration through profilaggrin-to-filaggrin processing. *EMBO Mol Med.* 3:320–
270 333.

271 Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard
272 P, Howes S, et al. 2000. Syncytin is a captive retroviral envelope protein
273 involved in human placental morphogenesis. *Nature* 403:785–789.

274 Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A,
275 Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for
276 phylogenetic inference in the genomic era. *Mol Biol Evol.* 37:1530–1534.

277 Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL,
278 Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2021. Pfam: The protein
279 families database in 2021. *Nucleic Acids Res.* 49:D412–D419.

280 Nakagawa S, Takahashi MU. 2016. gEVE: a genome-based endogenous viral element
281 database provides comprehensive viral protein-coding sequences in mammalian
282 genomes. *Database.* 2016:1–8.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 11:1650–1667.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 16:276–277.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 14:178–192.

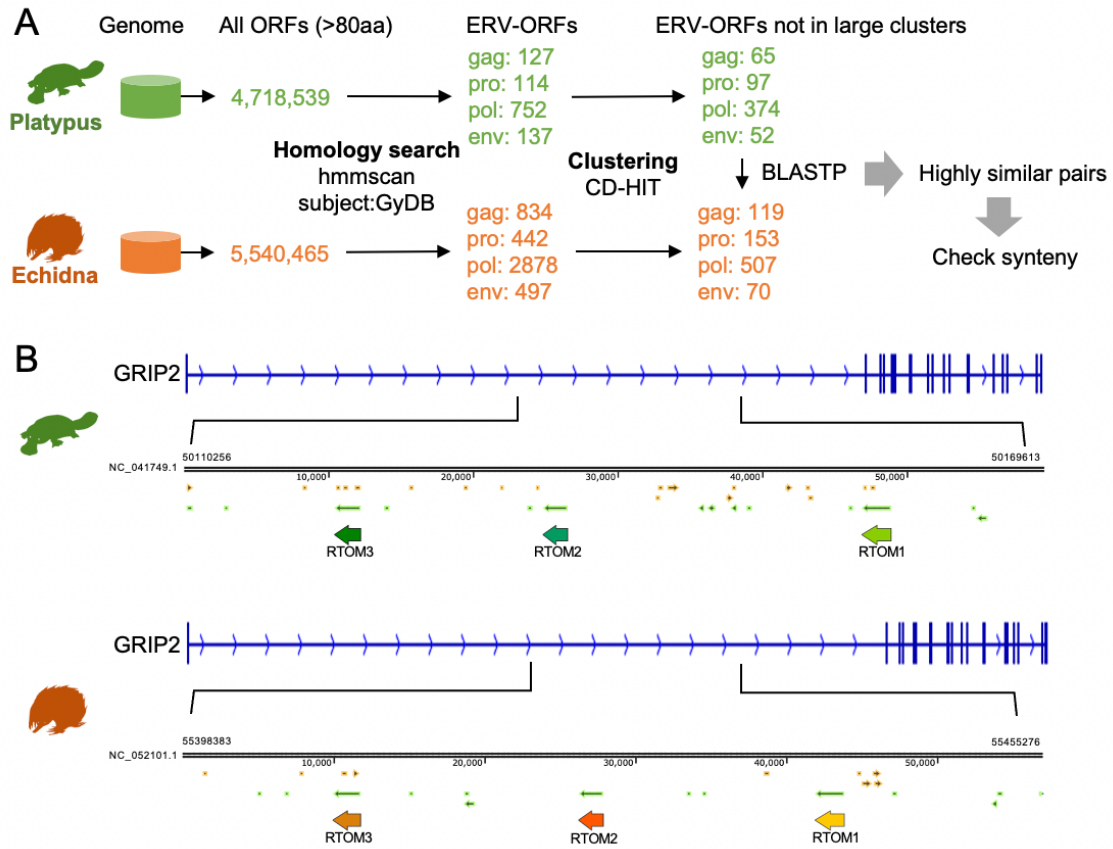
Ueda MT, Kryukov K, Mitsuhashi S, Mitsuhashi H, Imanishi T, Nakagawa S. 2020. Comprehensive genomic analysis reveals dynamic evolution of endogenous retroviruses that code for retroviral-like protein domains. *Mob DNA* 11:29.

Wang J, Han GZ. 2020. Frequent retroviral gene co-option during the evolution of vertebrates. *Mol Biol Evol.* 37:3232–3242.

Warren WC, Hillier LW, Marshall Graves JA, Birney E, Ponting CP, Grützner F, Belov K, Miller W, Clarke L, Chinwalla AT, et al. 2008. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* 453:175–183.

Zhou Y, Shearwin-Whyatt L, Li J, Song Z, Hayakawa T, Stevens D, Fenelon JC, Peel E, Cheng Y, Pajpach F, et al. 2021. Platypus and echidna genomes reveal mammalian biology and evolution. *Nature* 592:756–762.

302 **Figure legends**



303

304 FIG. 1. Identification of *RTOM1*, *2*, and *3*. (A) Schematic representation of the in silico

305 screening for conserved ERV-derived genes in platypus and echidna. (B) Genomic

306 context of *RTOM1*, *2*, and *3*.

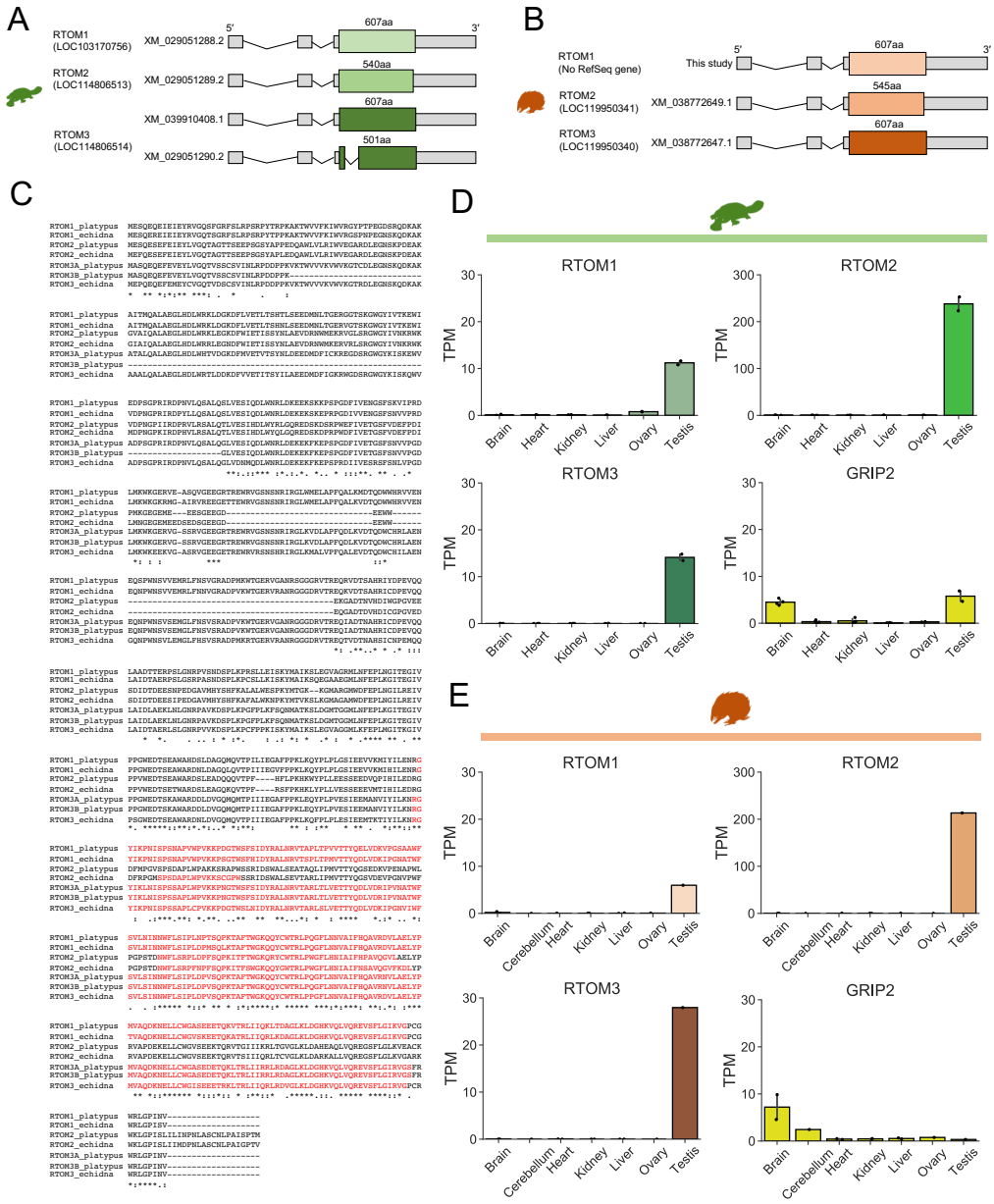


FIG. 2. Expression of *RTOM1*, *2*, and *3*. (A) Schematic representation of the RefSeq transcripts of the *RTOM* genes in platypus. (B) Schematic representation of the reconstructed *RTOM1* transcript and RefSeq transcripts of the *RTOM2* and *3* genes in echidna. (C) Multiple alignment of the amino acid sequences of RTOM proteins. The amino acid sequence of echidna RTOM1 was obtained from the genomic ORF. “RTOM3A_plasypus” and “RTOM3B_platypus” are protein isoforms derived from

314 “XM_039910408.1” and “XM_029051290.2,” respectively. The regions showing
315 similarity to the HMM of spumaretrovirus RT domain in GyDB are indicated in red. (D
316 and E) Tissue-specific expression of *RTOM* genes and *GRIP2* in (D) platypus and (E)
317 echidna. Normalized expression levels are presented as transcript per million (TPM).
318

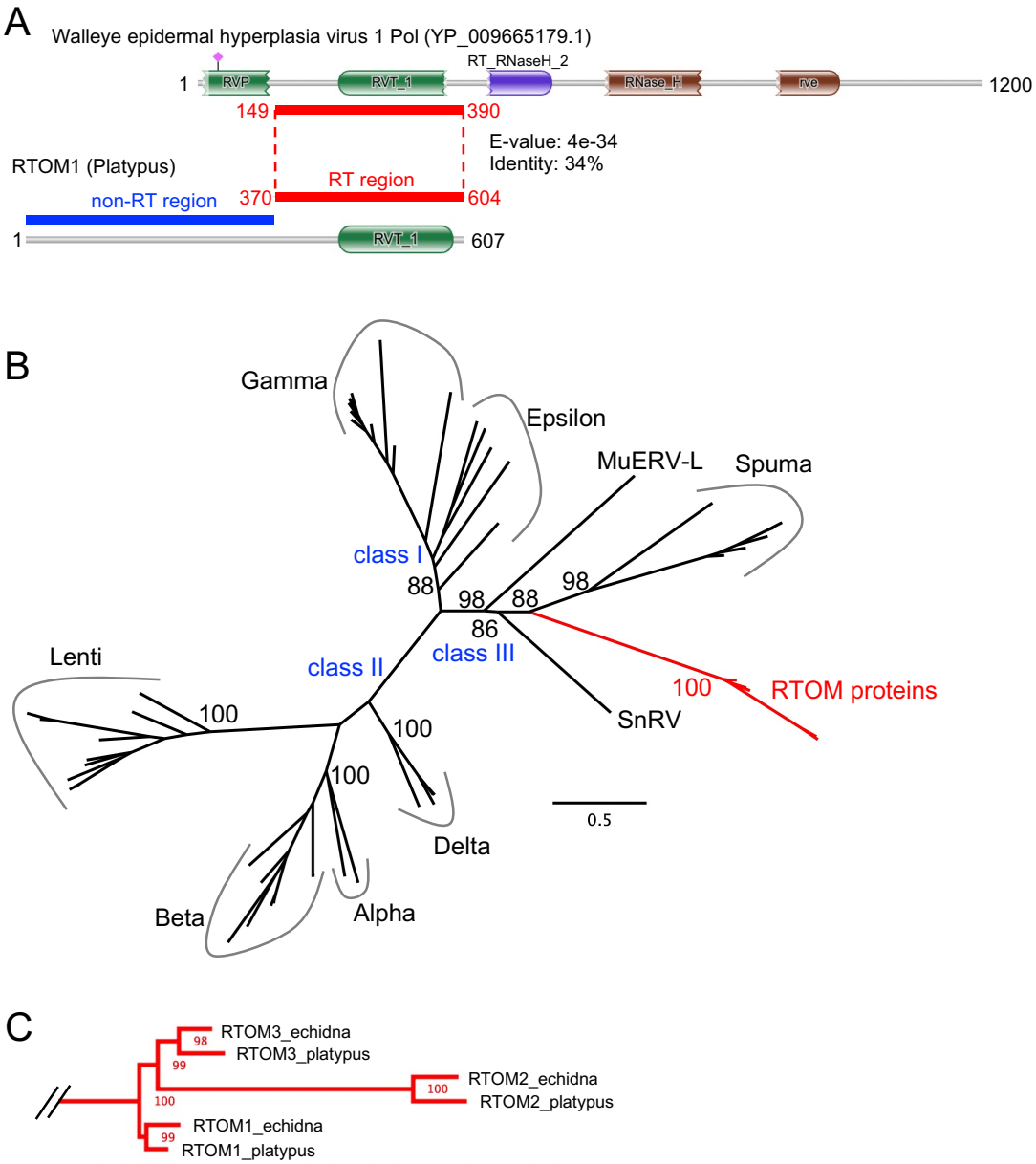
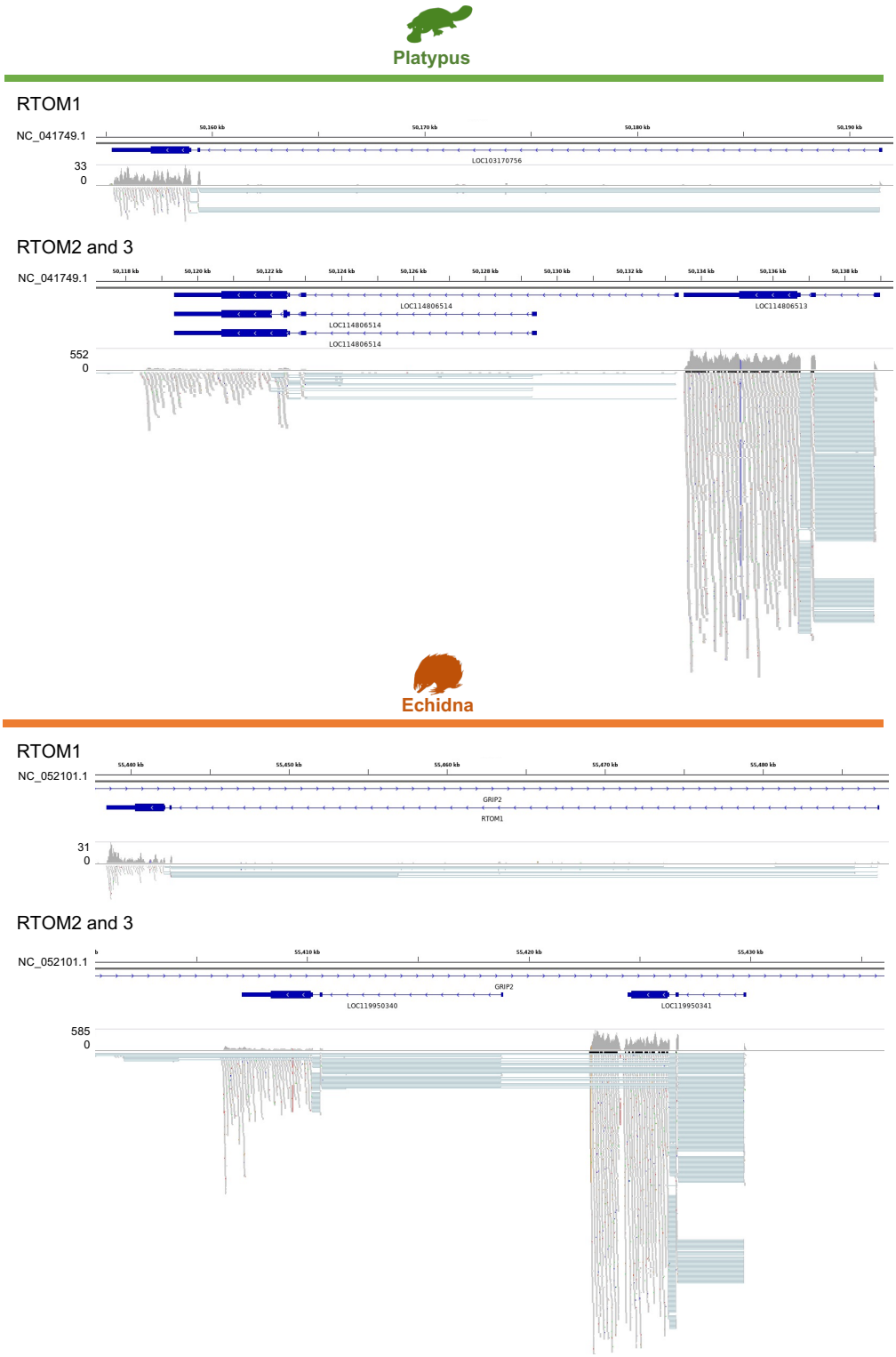


FIG. 3. Evolution of *RTOM1*, *2*, and *3*. (A) Comparison between platypus RTOM1 and retroviral Pol protein. Walleye epidermal hyperplasia virus 1 is represented as an example. A region showing similarity to the Pol protein by BLASTp was designated as "RT region." A region that did not show similarity 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 RTOM proteins and the retroviral Pol proteins in GyDB. The multiple alignment

326 is available in supplementary data S2. Ultrafast-bootstrap values obtained from 1000
327 times replication are shown in major branches. (C) Detailed representation of the clade
328 of the RTOM proteins shown in (B).
329



Supplementally FIG. S1. Screenshots of Interactive Genomic Viewer of RNA-seq reads on the *RTOM* genes. The transcript tracks in blue lines display the coordinates from the

333 RefSeq GTF files. Thick blue lines indicate the coding sequences. Since there is no
334 corresponding RefSeq transcript for echidna *RTOM1*, its gene coordinate was manually
335 added from assembled transcripts in this study (Materials and Methods).
336

Supplementally FIG. S2. Protein domains in RTOM1, 2, and 3. The domain search was conducted using hmmscan in HMMER web server with default options (https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan).