AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

A Missing Link between Retrotransposons and Retroviruses

Jianhua Wang,[a] **Guan-Zhu Han**[a]

[a]Jiangsu Key Laboratory for Microbes and Functional Genomics, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu, China

ABSTRACT The origin and deep evolution of retroviruses remain largely unclear. It has been proposed that retroviruses might have originated from a Ty3/Gypsy retrotransposon, but all known Ty3/Gypsy retrotransposons are only distantly related to retroviruses. Retroviruses and some plant Athila/Tat elements (within Ty3/Gypsy retrotransposons) independently evolved a dual RNase H domain and an *env*/*env*-like gene. Here, we reported the discovery of a novel lineage of retrotransposons, designated Odin retrotransposons, in the genomes of eight sea anemones (order Actinaria) within the Cnidaria phylum. Odin retrotransposons exhibited unique genome features, encoding a dual RNase H domain (like retroviruses) but no *env* gene (like most Ty3/Gypsy retrotransposons). Phylogenetic analyses based on reverse transcriptase showed that Odin retrotransposons formed a sister group to lokiretroviruses, and lokiretroviruses and Odin retrotransposons together were sister to canonical retroviruses. Moreover, phylogenetic analyses based on RNase H and integrase also supported the hypothesis that Odin retrotransposons were sisters to lokiretroviruses. Lokiretroviruses and canonical retroviruses did not form a monophyletic group, indicating that lokiretroviruses and canonical retroviruses might represent two distinct virus families. Taken together, the discovery of Odin retrotransposons narrowed down the evolutionary gaps between retrotransposons and canonical retroviruses and lokiretroviruses.

IMPORTANCE The origin of retroviruses remains largely unclear. In this study, we discovered a novel retrotransposon lineage, Odin retrotransposons, within the genomes of sea anemones (order Actinaria). In contrast to retroviruses and most retrotransposons, Odin retrotransposons encode a dual RNase H domain but no *env* gene. Phylogenetic analyses showed that Odin retrotransposons were sisters to lokiretroviruses, and lokiretroviruses and Odin retrotransposons were sisters to retroviruses, establishing an evolutionary framework to decipher the origin of retroviruses (canonical retroviruses and lokiretroviruses). Our results provided insights into the diversity and deep evolution of LTR retrotransposons closely related to retroviruses.

KEYWORDS Comparative genomics, evolution, phylogenetic analysis, retrotransposons, retroviruses

Retroviruses (the *Retroviridae* family) infect a wide range of vertebrates, and their replication requires reverse transcription and integration into host genomes (1–3). While retroviruses usually infect somatic cells (4–7), they occasionally infect germline cells and become integrated into the genome and may be vertically inherited, forming so-called endogenous retroviruses (ERVs) (1–3). Canonical exogenous retroviruses have been typically classified into two subfamilies, namely, *Orthoretrovirinae* (including alpharetroviruses, betaretroviruses, gammaretroviruses, deltaretroviruses, epsilonretroviruses, and lentiviruses) and *Spumaretrovirinae* (foamy viruses) (8). Based on their relationships with exogenous retroviruses, ERVs are traditionally classified into Class I (closely related to gammaretroviruses), Class II (closely related to betaretroviruses), and Class III (closely related to foamy viruses) (4, 9). However, the classification system of exogenous and endogenous retroviruses has not been well incorporated (7, 9).

Recently, a putatively new subfamily of retroviruses, designated lokiretroviruses, has been discovered in the genomes of vertebrates, including lampreys, fishes, amphibians, and reptiles

Invited Editor Alex Hayward, University of Exeter

Editor Vaughn S. Cooper, University of Pittsburgh

Address correspondence to Guan-Zhu Han, guanzhu@njnu.edu.cn.

The authors declare no conflict of interest.

Received 21 January 2022
Accepted 17 February 2022
Published 15 March 2022

(10). Lokiretroviruses display some unique genome features: (i) Like canonical retroviruses, lokiretroviruses encode a dual RNase H (RH) domain. They acquired a new RH domain, and the preexisting RH domain degenerated to a tether domain (10–13). (ii) Lokiretroviruses encode Env proteins that share detectable sequence similarity with fusion glycoproteins of viruses within the Mononegavirales order (nonsegmented negative-sense single-stranded RNA viruses), but not canonical retroviruses (10). Phylogenetic analyses based on reverse transcriptase (RT) proteins suggest that lokiretroviruses are sister to all the sampled canonical retroviruses, and thus lokiretrovirus was tentatively classified as a novel subfamily within the family *Retroviridae* (10). Thereafter, we used retroviruses to refer to canonical retroviruses and lokiretroviruses, unless otherwise specified. The discovery of lokiretroviruses corroborates the complex evolutionary history of retroviruses (10).

Five reverse transcribing viruses, namely, *Retroviridae*, *Metaviridae* (Ty3/Gypsy retrotransposons), *Pseudoviridae* (Ty1/Copia retrotransposons), *Belpaoviridae* (Bel-Pao retrotransposons), and *Caulimoviridae* (plant pararetroviruses), have recently been unified into the viral order *Ortervirales* (14). Members within the *Ortervirales* order are thought to have originated from a common ancestor (14, 15). Phylogenetic analyses based on RT show that retroviruses are more closely related to the Ty3/Gypsy retrotransposons within the *Ortervirales* order (10, 14–17). Thus, it has been hypothesized that retroviruses might have originated from a Ty3/Gypsy retrotransposon through acquiring an *env* gene (18–20). However, all the known Ty3/Gypsy retrotransposons are only distantly related to retroviruses. It remains largely unclear how retroviruses originated (20).

In this study, we performed systematic mining of retrotransposons that are closely related to retroviruses within 3,624 animal genomes. Intriguingly, we discovered a novel retrotransposon lineage closely related to retroviruses in the genomes of eight sea anemones (order Actinaria) in the phylum Cnidaria. The newly discovered retrotransposons exhibited unique genome features. Evolutionary analyses of the newly discovered retrotransposons provided insights into the diversity and deep evolution of LTR retrotransposons closely related to retroviruses.

RESULTS

The discovery of Odin retrotransposons. To investigate the origin of retroviruses, we used a similarity search and phylogenetic analyses combined approach to screen retroelements that are closely related to retroviruses within 3,624 animal genomes (1,768 Vertebrata, 1,710 Protostomia, 74 Cnidaria, 27 Echinodermata, 19 Tunicata, 8 Cephalochordata, 8 Porifera, 4 Ctenophora, 2 Hemichordata, 2 Xenacoelomorpha, and 2 Placozoa) (Fig. 1A; Table S1 and S2) retrieved from NCBI. Endogenous canonical retroviruses and lokiretroviruses have been only identified within the genomes of vertebrates (Fig. 1B) (7, 10). Intriguingly, we identified a novel lineage of retrotransposons in the genomes of eight sea anemones (order Actinaria) within the Cnidaria phylum (Fig. 1A to C and Fig. 2A). We designated the retrotransposon lineage Odin retrotransposons following the name of Odin in Norse mythology, the blood brother of Loki after whom lokiretroviruses was named (10). The copy numbers of Odin retrotransposons were generally low within cnidaria genomes, ranging from one in *Exaiptasia pallida* to six in *Heteractis magnifica* (Table S3). Moreover, several Odin retrotransposons integrated into host genomes in recent time (from 0 to 4.66 million years ago; Table S3), suggesting that some Odin retrotransposons might still be active in their host genomes.

Retrovirus genomes encode dual RH domains and *env* genes, whereas most long terminal repeat (LTR) retrotransposons do not encode dual RH domains or *env* genes [with a few exceptions, such as some land plant Athila/Tat elements within Ty3/Gypsy retrotransposons; (12, 13)]. Retroviruses and Athila/Tat elements independently evolved a dual RNase H domain and an *env*/*env*-like gene (12, 13). Odin retrotransposons encoded two putative genes that were common to LTR retrotransposons, namely, *gag* and *pol*, flanked by two LTRs (Fig. 3A). No *env*-like gene was predicted within Odin retrotransposons. Odin retrotransposon Pol proteins comprised four domains, including protease (PR), RT, RH, and integrase (IN) (Fig. 3A). Interestingly, like retroviruses and unlike most LTR retrotransposons, Odin retrotransposons encoded a dual RH domain that consisted of a tether domain derived

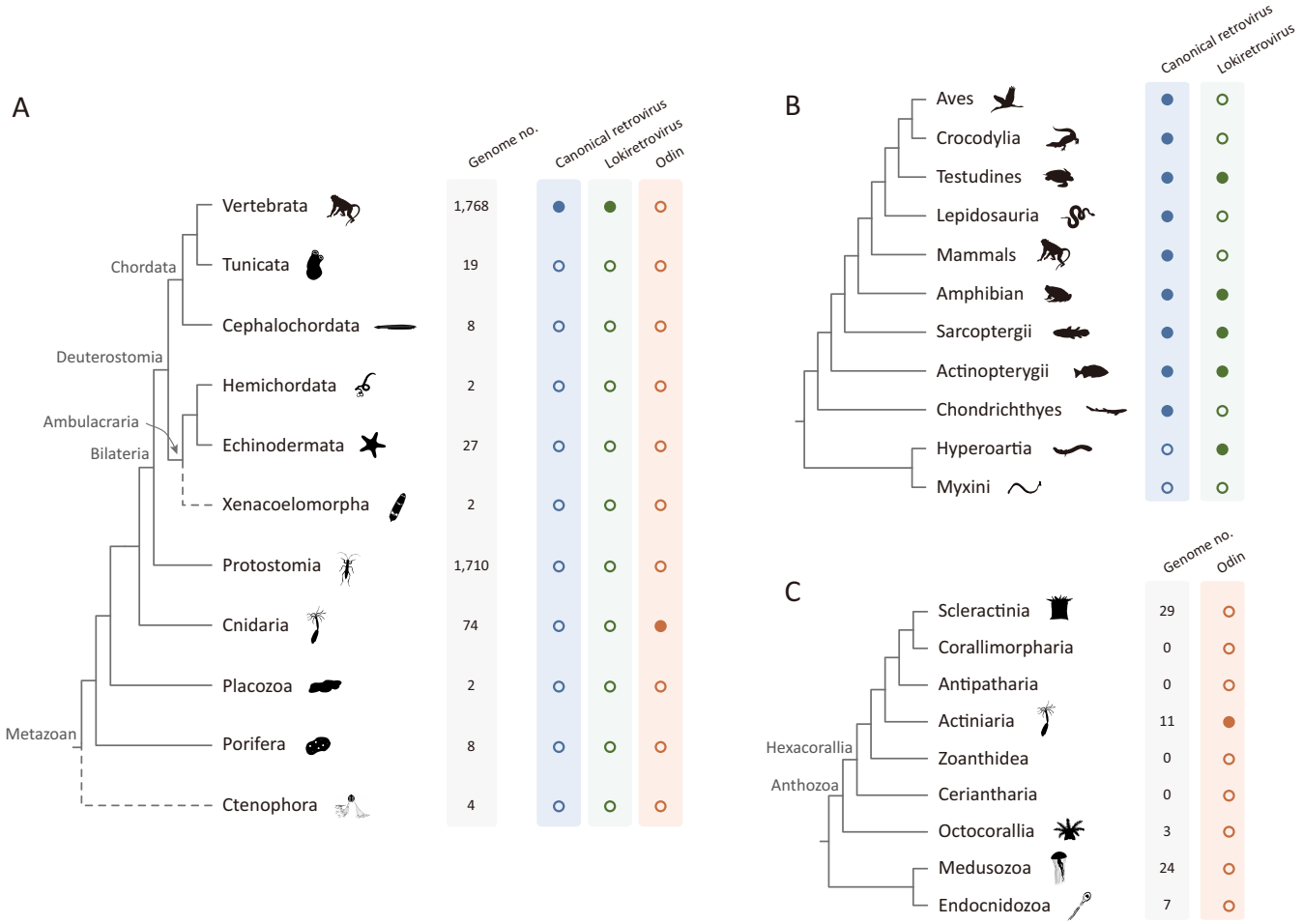

FIG 1 Host distribution of Odin retrotransposons and retroviruses. (A) Distribution of Odin retrotransposons, lokiretroviruses, and canonical retroviruses in metazoans. (B) Distribution of lokiretroviruses and canonical retroviruses in vertebrates. (C) Distribution of Odin retrotransposons in Anthozoa. Phylogenetic relationships of metazoans are based on the literature (34–38). The dashed lines indicate phylogenetic uncertainty. Genome no. represents the number of genomes used in this study. The filled blue, green, and orange circles represent the presence of canonical retroviruses, lokiretroviruses, and Odin retrotransposons in the corresponding animal groups, respectively. The open blue, green, and orange circles represent the absence of canonical retroviruses, lokiretroviruses, and Odin retrotransposons in the corresponding animal groups, respectively.

from the degenerated Ty3/Gypsy retrotransposon RH domain and a newly acquired RH domain (10–13). The detectable structural similarity was found between the tether domains of Odin retrotransposon and canonical retroviruses (for example, Odin retrotransposon versus human immunodeficiency virus type 1 [HIV-1]: probability = 99.92%, *E* value = 1.6×10^{-31}, identities = 6%; Odin retrotransposon versus murine leukemia virus [MuLV]: probability = 100%, *E* value = 2.6×10^{-42}, identities = 22%) (Fig. 3B). Moreover, the tether domain of Odin retrotransposon shared detectable structural similarity with the RH domain of the Ty3 retrotransposon (probability = 99.94%, *E* value = 4.7×10^{-32}, identities = 27%) (Fig. 3B). Taken together, our results showed that Odin retrotransposons exhibited unique genome features, encoding a dual RNase H domain (like retroviruses) but no *env* gene (like most of Ty3/Gypsy retrotransposons).

Odin retrotransposons are sister to lokiretroviruses. To explore the evolutionary relationship among Odin retrotransposons, retroviruses, and other *Ortervirales* members, we performed phylogenetic analyses based on RT protein alignments generated by two different methods (MAFFT with the L-INS-I strategy [align-Ma] and PROMAL3D with the default parameters [align-3D]) using two tree reconstruction algorithms (maximum likelihood and Bayesian inference) (Table S2) and obtained four largely consistent phylogenies. Phylogenetic analyses showed that Odin retrotransposons formed a sister group of lokiretroviruses with robust support values (ultrafast bootstrap approximation

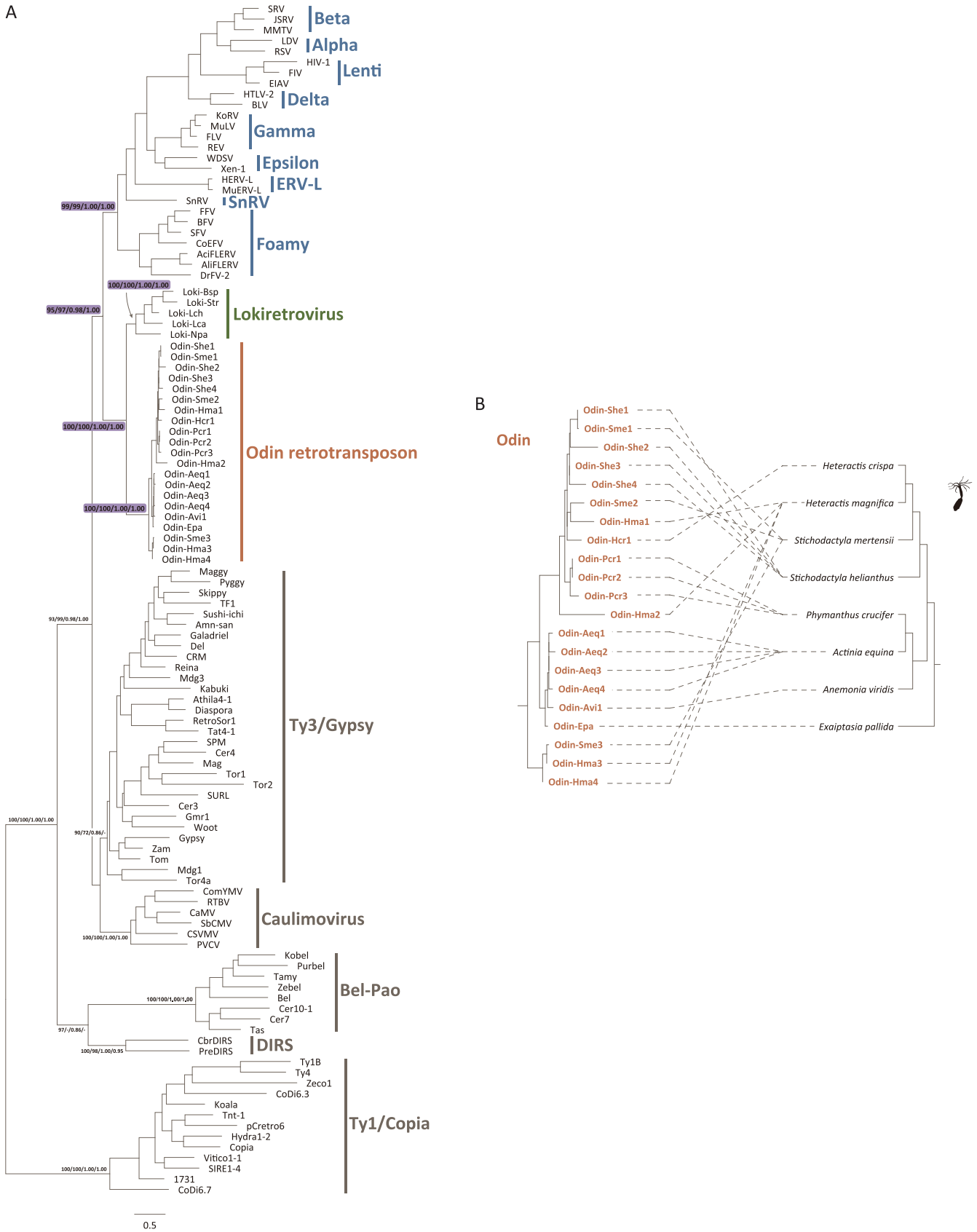

FIG 2 Phylogenetic relationships among Odin retrotransposons, representative canonical retroviruses, lokiretroviruses, and LTR retrotransposons. (A) Phylogenetic trees were reconstructed based on RT proteins of Odin retrotransposons, representative canonical retroviruses, lokiretroviruses, and LTR retrotransposons. (B) Comparison of Odin retrotransposon and host phylogenies. The left is the Odin retrotransposon phylogeny, whereas the right is the host phylogeny based on the literature (39).

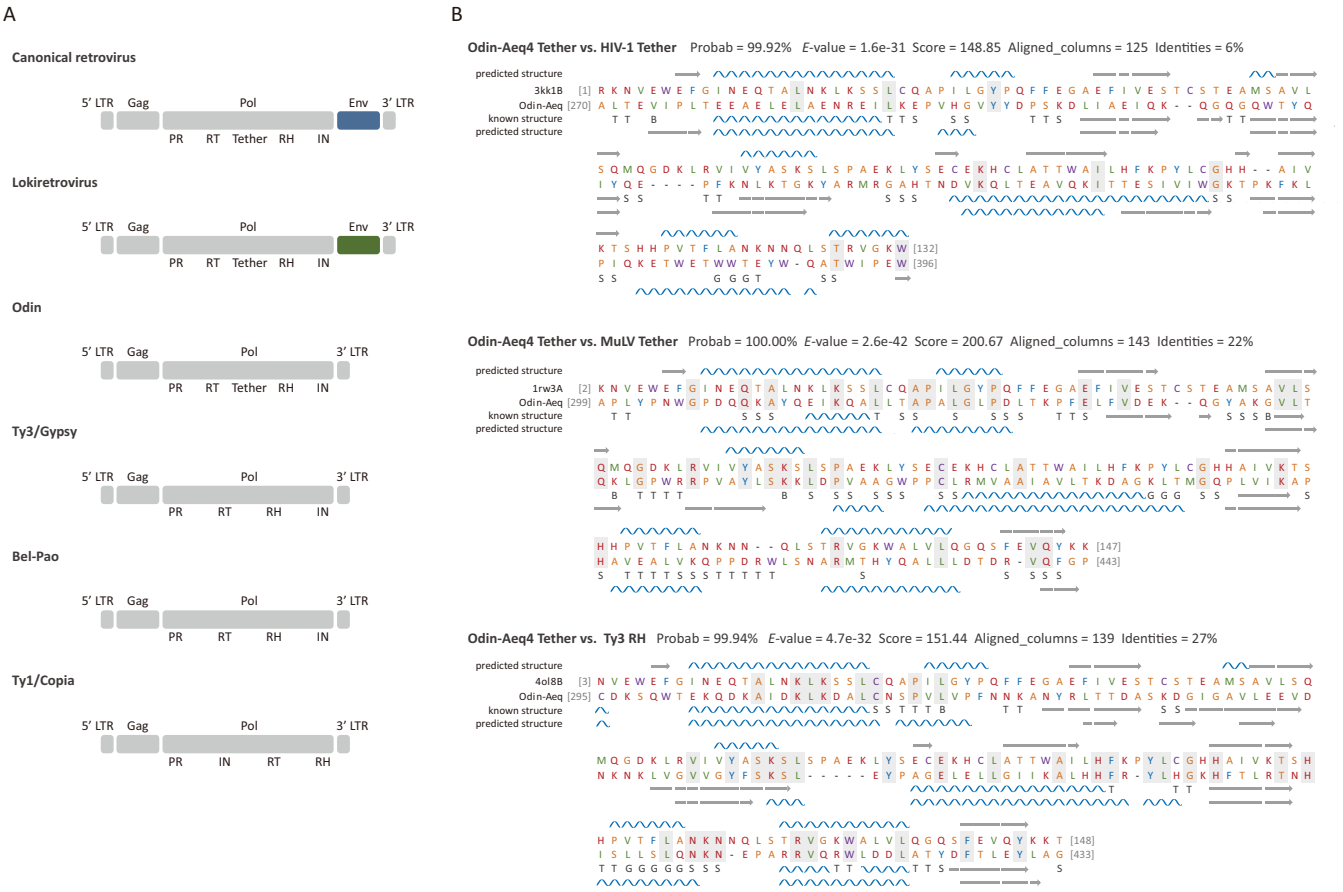

FIG 3 Domain architectures and secondary structure of retroviruses and retrotransposons. (A) Comparison of domain architectures of canonical retroviruses, lokiretroviruses, Odin retrotransposons, and other retrotransposons. (B) Comparison of secondary structure between the tether domain of Odin-Aeq4 and that of HIV-1 (PDB accession no.: 3KK1.B) and MuLV (PDB accession no.: 1RW3.A), and between the tether domain of Odin-Aeq4 and the RH domain of Ty3 (PDB accession no.: 4OL8.A). The blue helices and orange arrows represent α-helices and β-strands, respectively. "T," "S," and "G" characters represent the following: T = hydrogen-bonded turn, S = bend, and G = 310 helix.

[UFBoot] = 100% for both alignments and Bayesian posterior probability [BPP] = 1.00 for both alignments) (Fig. 2A). Canonical retroviruses (*Orthoretrovirinae* and *Spumaretrovirinae*) did not form a monophyletic group with lokiretroviruses. Rather, Odin retrotransposons and lokiretroviruses formed a sister group to canonical retroviruses (UFBoot = 95% and 97% for align-Ma and align-3D, respectively; BPP = 0.98 and 1.00 for align-Ma and align-3D, respectively) (Fig. 2A). Taken together, our results suggested that Odin retrotransposons were sister to lokiretroviruses and the closest known retrotransposon relatives of retroviruses (canonical retroviruses and lokiretroviruses).

Frequent host switching of Odin retrotransposons. Initially, we found that the phylogeny of Odin retrotransposons based on RT protein sequences was generally inconsistent with host phylogeny (Fig. 2B). We further used an event-based approach to compare the congruence of phylogenies between Odin retrotransposons and their hosts at the species level and the family level. We found significant incongruence between phylogenies of Odin retrotransposons and their hosts at the species level ($P > 0.05$; Table 1) and at the family level ($P > 0.05$; Table 1). These results suggested that Odin retrotransposons mainly evolved through frequent host switching (horizontal transfers).

The evolutionary history of lokiretrovirus RH and IN domains. To further explore the evolutionary history of Odin retrotransposons and retroviruses, we performed phylogenetic analyses based on their RH and IN domains. Broadly consistent with previous studies (10, 13), both RH and IN domains of canonical retroviruses clustered into multiple groups, suggesting that canonical retroviruses frequently replaced their RH and IN

TABLE 1 Test for congruence of phylogenies between Odin retrotransposons and their hosts

Datasets	Event costs[a]	Total cost	No. of events					*P* value[b]
			Cospeciation	Duplication	Duplication and host switching	Loss	Failure to diverge	
Species	0, 1, 2, 1, 1	27	3	8	9	1	0	$P > 0.05$
Species	0, 1, 1, 2, 0	18	2	6	12	0	0	$P > 0.05$
Species	−1, 0, 0, 0, 0	−5	5	6	9	9	0	$P > 0.05$
Family	0, 1, 2, 1, 1	21	2	15	3	0	0	$P > 0.05$
Family	0, 1, 1, 2, 0	18	2	15	3	0	0	$P > 0.05$
Family	−1, 0, 0, 0, 0	−2	2	15	3	0	0	$P > 0.05$

[a]Event cost schemes are for cospeciation, duplication, duplication with host switch, loss, and failure to diverge, respectively.
[b]*P* value represents statistical analysis results by using the method of random parasite tree with a sample size of 500.

domains during their evolutionary history (Fig. 4) (10). Consistent with phylogenetic analyses of the RT domain, we found that Odin retrotransposons were sister to lokiretroviruses for both RH and IN domains with robust support values (RH domain: UFBoot = 100% for both alignments, BPP = 1.00 for both alignments; IN domain: UFBoot = 98% and 100% for align-Ma and align-3D, respectively, and BPP = 0.83 and 1.00 for align-Ma and align-3D, respectively) (Fig. 4). It should be noted that, because phylogenetic analyses of RH and IN are notoriously problematic (10), we could not infer a robust evolutionary relationship among Odin retrotransposons, lokiretroviruses, and canonical retroviruses for RH or IN domains. Nevertheless, our phylogenetic analyses of RH and IN domains further supported that Odin retrotransposons were sister to lokiretroviruses.

DISCUSSION

In this study, we discovered a novel lineage of retrotransposons, referred to as Odin retrotransposons, within eight sea anemones (order Actiniaria, phylum Cnidaria). Odin retrotransposons exhibited unique genome features, encoding a dual RNase H domain (like retroviruses) but no Envelope protein (like most of Ty3/Gypsy retrotransposons). Our phylogenetic analyses showed that Odin retrotransposons formed a sister group to lokiretroviruses, and Odin retrotransposons and lokiretroviruses were sisters to canonical retroviruses. Lokiretroviruses and canonical retroviruses did not form a monophyletic group. Therefore, Odin retrotransposons were the closest known retrotransposon relatives to retroviruses (canonical retroviruses and lokiretroviruses). Retroviruses have long been thought to have originated from an ancient Ty3/Gypsy retrotransposon (15, 16, 20), but the sampled Ty3/Gypsy retrotransposons were only distantly related to retroviruses. The discovery of Odin retrotransposons narrowed down the evolutionary gap between Ty3/Gypsy retrotransposons and retroviruses. Odin retrotransposons might represent the modern descendants of those long-sought-after Ty3/Gypsy retrotransposons.

Ty3/Gypsy retrotransposons typically encode two common genes (*gag* and *pol*), whereas both canonical retroviruses and lokiretroviruses encode an additional gene, *env*, besides *gag* and *pol* genes (18). Moreover, most of Ty3/Gypsy retrotransposon (with land plant Athila/Tat retrotransposons as the exception) Pol proteins comprise PR, RT, RH, and IN domains, whereas both canonical retroviruses and lokiretroviruses encode the dual RH domain. The preexisting RH domain degenerated to a tether domain and a new RH domain was acquired (10–13). Interestingly, Odin retrotransposons also possess the dual RH domain with a degraded RH domain (the tether domain) but no Env proteins. Therefore, the genome architecture of Odin retrotransposons might represent an intermediate formed between Ty3/Gypsy retrotransposons and canonical retroviruses/lokiretroviruses.

The sequence identity between the tether domains of canonical retroviruses and the RH domains was too low to be used for phylogenetic analyses, but the tether domains of Odin retrotransposons and lokiretroviruses shared detectable structural similarity with the tether domains of canonical retroviruses (HIV-1 and MuLV) and the RH domain of Ty3/Gypsy retrotransposons (the Ty3 retrotransposon) (10). Moreover, phylogenetic analyses of RT proteins showed that Odin retrotransposons, lokiretroviruses, and canonical retroviruses clustered together. Therefore, we inferred that the degradation of the preexisting RH domain occurred

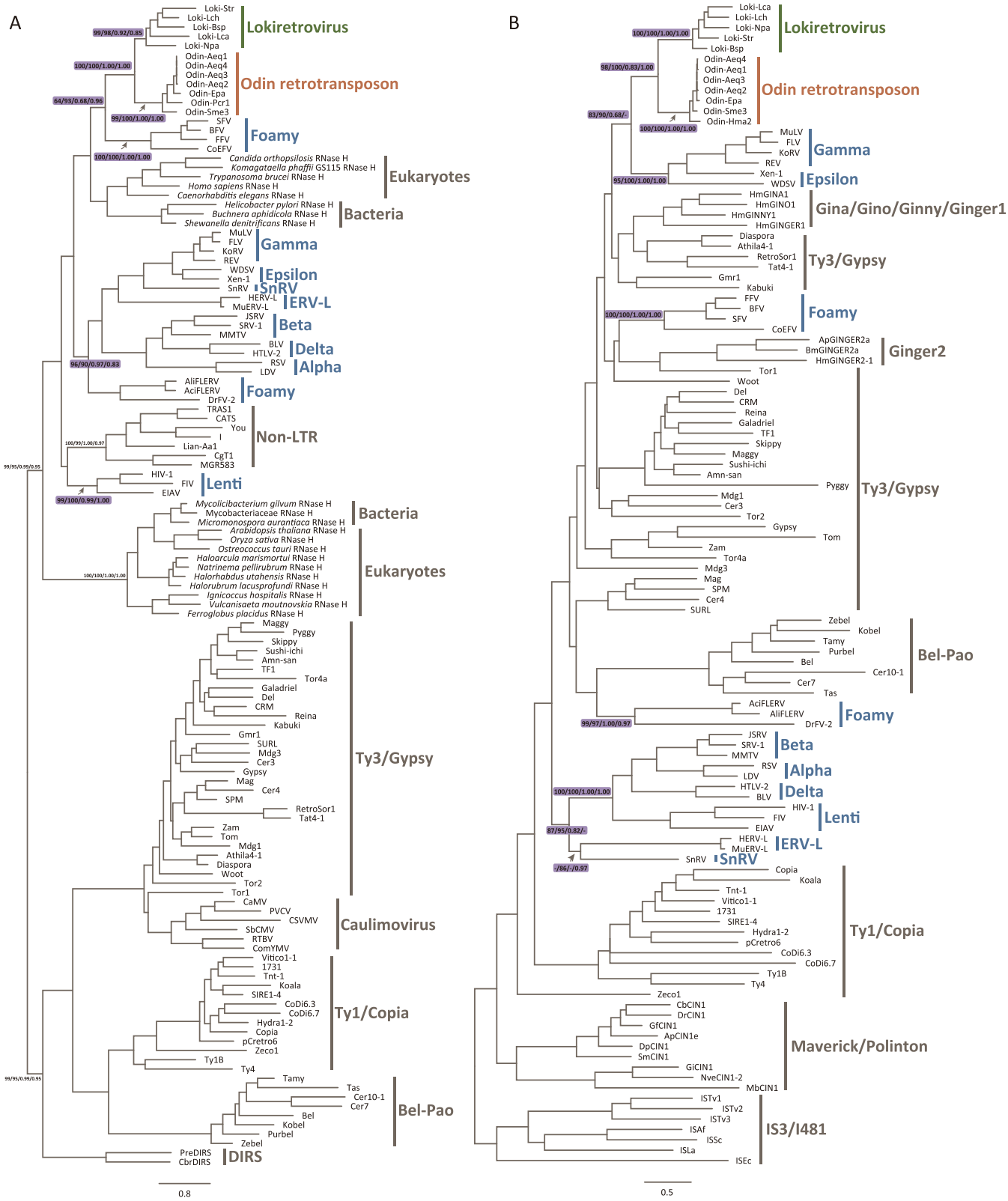

FIG 4 Phylogenetic trees of RH and IN domains. (A) Phylogenetic tree of RH domains. (B) Phylogenetic tree of IN domains. The support values are listed in the order of UFBoot for align-Ma/UFBoot for align-3D/BPP for align-Ma/BPP for align-3D.

in the most recent common ancestor (MRCA) of Odin retrotransposons and retroviruses. For canonical retroviruses, the newly acquired RH domains did not cluster together, indicating canonical retroviruses replaced their RH domains multiple times (10, 13).

Our phylogenetic analyses provided a crucial framework for investigating the origin and evolution of retroviruses (canonical retroviruses and lokiretroviruses). The MRCA of Odin retrotransposons and retroviruses was likely to be an LTR retrotransposon with a dual RH domain. Env proteins of canonical retroviruses and lokiretroviruses did not share detectable similarities, but instead, lokiretrovirus Env proteins shared detectable similarities with fusion glycoproteins of viruses within the Mononegavirales, suggesting that Env proteins of canonical retroviruses and lokiretroviruses were likely to be of different origins (10). Based on currently available information, at least four different evolutionary scenarios could be conceived to account for the origin of retroviruses. Independent origins of canonical retroviruses and lokiretroviruses (Fig. 5A and I). Canonical retroviruses and lokiretroviruses originated independently by acquiring Env proteins from different sources. The ancestor of lokiretroviruses might have acquired an Env protein from negative-sense single-stranded viruses (10). Odin retrotransposons might represent one of the modern descendants of these retrotransposon ancestors of retroviruses. The MRCA of Odin retrotransposons and retroviruses acquired an *env* gene. The *env* gene was replaced along with the evolution of canonical retroviruses, lokiretroviruses, or the MRCA of lokiretroviruses and Odin retrotransposons. The *env* gene was then lost in Odin retrotransposons (Fig. 5A, II). Canonical retroviruses and the MRCA of lokiretroviruses and Odin retrotransposons independently acquired *env* genes, and the *env* gene was lost during the evolutionary course of Odin retrotransposons (Fig. 5A, III). The MRCA of Odin retrotransposons and retroviruses acquired an *env* gene. The *env* gene was lost in the MRCA of Odin retrotransposons and lokiretroviruses. But lokiretroviruses acquired a new *env* gene during their evolutionary course (Fig. 5A, IV). Among these evolutionary scenarios, scenario I (two gain events) was more parsimonious than the other scenarios (at least three steps). Therefore, we prefer the hypothesis that canonical retroviruses and lokiretroviruses originated independently by acquiring Env proteins from different sources (Fig. 5B). If so, more Odin-like retrotransposons with a dual RH domain but without Env proteins (indicated by dashed lines in Fig. 5B) await to be discovered possibly within the genomes from metazoan groups that are underrepresented in genome sequencing projects (outside Vertebrata and Protostomia). Moreover, given the small sample of Odin retrotransposons currently identified, a larger sample is required to further corroborate the lack of an *env* gene in the group.

Both canonical retroviruses and lokiretroviruses have been thought to infect vertebrates exclusively and widely. Odin retrotransposons were discovered in only eight sea anemones (only 11 anemone genomes were screened) with low copy numbers [ranging from one in *Exaiptasia pallida* to six in *Heteractis magnifica* (Table S3)]. Compared with Odin retrotransposons, canonical retroviruses and lokiretroviruses appear to be more widespread. The potential for infectivity (i.e., the presence of an *env* gene) of retroviruses and lokiretrovirus might entail the likelihood of host range expansion. However, evolutionary gaps remain between Odin retrotransposons of cnidaria and canonical retroviruses/lokiretroviruses of vertebrates. Two possibilities might account for the gaps: (i) horizontal transfer of Odin-like retrotransposons occurred between cnidaria and vertebrates; (ii) compared with many Vertebrata genomes (1,768) and Protostomia genomes (1,710), only 146 genomes from metazoans outside vertebrates and protostomia have been sequenced, and many more Odin or Odin-like retrotransposons (illustrated using dashed lines in Fig. 5B) might exist within metazoans.

In our previous study (10), lokiretroviruses were classified as a subfamily of retroviruses. However, our phylogenetic analyses show that lokiretroviruses and canonical retroviruses do not form a monophyletic group after adding the newly discovered Odin retrotransposons. Moreover, canonical retroviruses and lokiretroviruses might have acquired their Env proteins independently. Therefore, lokiretrovirus might represent a misnomer, and we think it is necessary to rename lokiretrovirus to lokiortervirus, which reflects that it belongs to the viral order *Ortervirales*. Unlike most Ty3/Gypsy retrotransposons (*Metaviridae*), Odin retrotransposons encode the dual RH domain, and Odin retrotransposons do not cluster within the diversity of Ty3/Gypsy retrotransposons. Taken together, we propose that lokiorterviruses and Odin

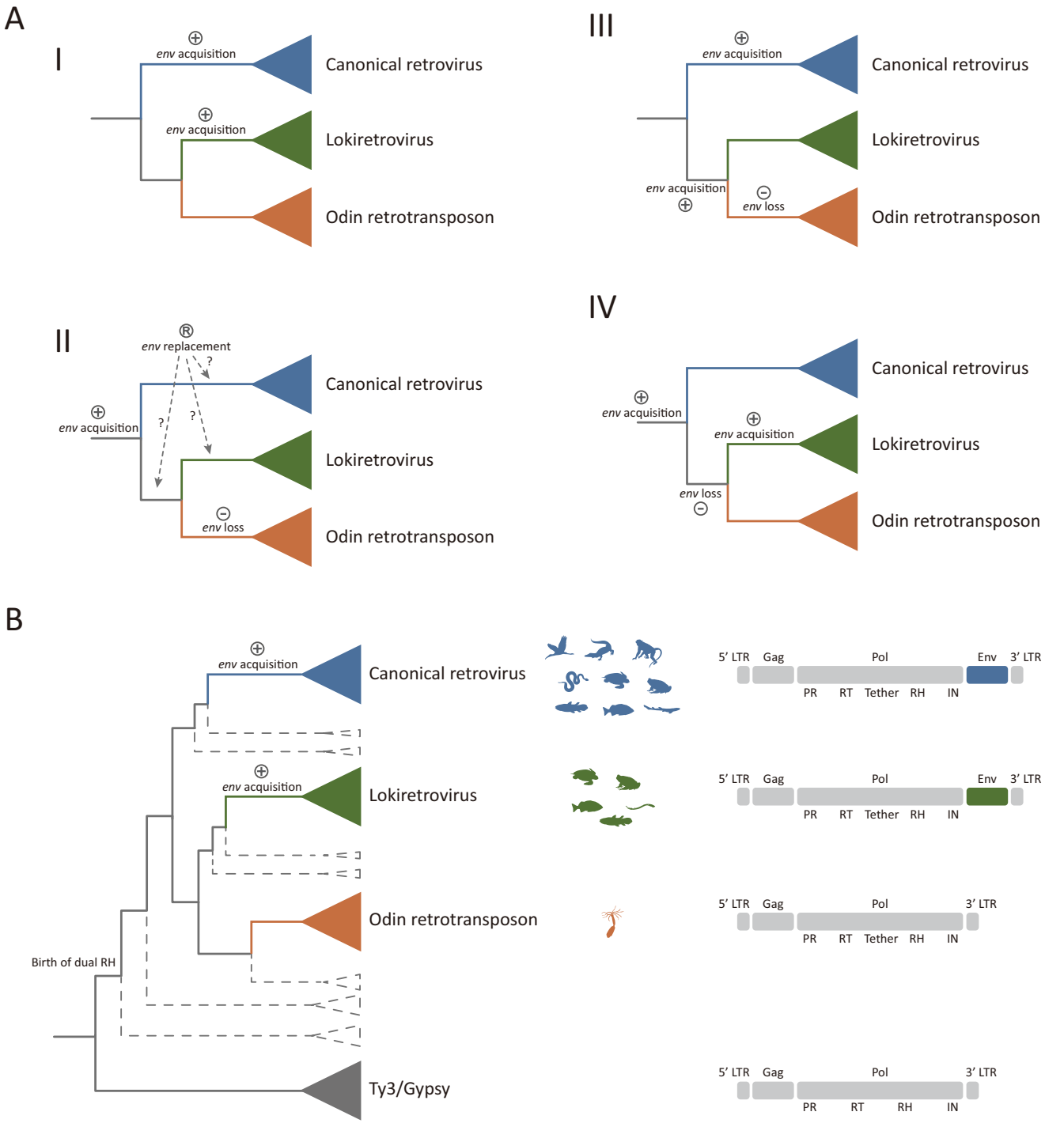

FIG 5 Evolutionary scenarios for the origin of retroviruses. (A) Four possible scenarios of the origin of retroviruses. Scenario I: Canonical retroviruses and lokiretroviruses acquired Env proteins independently from different sources. Scenario II: The MRCA of Odin retrotransposons and retroviruses acquired an *env* gene. Canonical retroviruses, lokiretroviruses, or the MRCA of lokiretroviruses and Odin retrotransposons replaced their *env* genes during the evolutionary course. Odin retrotransposons then lost the *env* gene. Scenario III: Canonical retroviruses and the MRCA of lokiretroviruses and Odin retrotransposons acquired *env* genes independently and then Odin retrotransposons lost the *env* gene. Scenario IV: The MRCA of Odin retrotransposons and retroviruses acquired an *env* gene, and the MRCA of Odin retrotransposons and lokiretroviruses lost the *env* gene. Lokiretroviruses then acquired a new *env* gene. (B) Model for the retrovirus origin. Canonical retroviruses and lokiretroviruses originated independently through the acquisitions of Env proteins from different sources. Dashed lines represent Odin-like retrotransposons that remain to be identified.

retrotransposons can be tentatively classified into two novel viral families (*Lokiorterviridae* and *Odinorterviridae*) within the order *Ortervirales*.

MATERIALS AND METHODS

The discovery of Odin retrotransposons. We used a similarity search and phylogenetic analyses combined approach (10) to mine retroelements that are closely related to retroviruses. First, we used the tBLASTn algorithm to screen homologs of retrovirus RT proteins within 3,624 animal genomes available in NCBI (including 1,768 Vertebrata, 1,710 Protostomia, 74 Cnidaria, 27 Echinodermata, 19 Tunicata, 8 Cephalochordata, 8 Porifera, 4 Ctenophora, 2 Hemichordata, 2 Xenacoelomorpha, and 2 Placozoa) using 10 representative canonical retrovirus and lokiretrovirus RT proteins as queries with an *E* cutoff value of 10^{-5} (Fig. 1; Table S1). We then performed phylogenetic analyses of significant hits with the length of >50 amino acids (aa) and RT proteins of representative canonical retroviruses, lokiretroviruses, and retrotransposons (21). RT protein sequences were aligned using MAFFT v7.475 (22). The initial large-scale phylogenetic analyses were performed using an approximate maximum likelihood method implemented in FastTree 2.1.10 (23). Some significant hits from several sea anemones, referred to as Odin retrotransposons form a sister group to lokiretroviruses. To further confirm the distribution of Odin retrotransposons, we performed a second round of similarity search with an Odin retrotransposon RT protein of *Actinia equine* (NCBI accession no. WHPX01000927.1, from 148,958 to 149,728; referred to as Odin-Aeq4) as the query and an *E* cutoff value of 10^{-5}. Phylogenetic analyses were performed using the approaches described above. Significant hits that cluster with Odin retrotransposons were retrieved for further study.

Genome structure reconstruction and secondary structure prediction. We bidirectionally extended the retrieved Odin retrotransposon significant hits and predicted the domain architecture using the conserved domain (CD) search with default parameters (24). LTR_Finder was used to predict the 5′- and 3′-LTRs with default parameters (25). The Phyre2 web was used to compare the secondary structure between the tether domain of an Odin retrotransposon within *Actinia equine* (namely, Odin-Aeq4), the tether domains of HIV-1 and MuLV, and between the tether domain of Odin-Aeq4 and the RH domain of the Ty3 retrotransposon (26).

Phylogenetic analyses. To explore the relationship among Odin retrotransposons, canonical retroviruses, lokiretroviruses, and LTR retrotransposons, their RT, RH, and IN protein sequences were used to perform phylogenetic analyses (Table S2 and S3). Sequences were aligned using MAFFT v7.475 with the L-INS-I strategy and PROMALS3D, an alignment tool based on enhanced information from database searches, secondary structure prediction, and 3D structures, with the default parameters, and ambiguous regions were manually removed (Data Set S1 contains original alignments, and Data sets S2-S7 are alignments manually refined) (22, 27). The length of the RT alignments is 252 aa and 326 aa for align-Ma and align-3D, respectively (Data Set S2 and S3). The length of the RH alignments is 198 aa and 191 aa for align-Ma and align-3D, respectively (Data Set S4 and S5). The length of the IN alignments is 270 aa and 225 aa for align-Ma and align-3D, respectively (Data Set S6 and S7). A maximum likelihood approach implemented in IQ-Tree 2 was used to perform phylogenetic analyses (28). The best-fit models were estimated by Model Finder (29). The best-fit model for each alignment is as follows: LG+R6 for multiple sequence alignments of RT and RH proteins generated by MAFFT, LG+R5 for multiple sequence alignments of RT and RH proteins generated by PROMALS3D, and LG+F+R6 for both multiple sequence alignments of IN proteins. Node supports were assessed using ultrafast bootstrap approximation with 1,000 replicates (30). Moreover, we also performed phylogenetic analyses using a Bayesian method implemented in MrBayes 3.2.7a (31). The best-fit models were selected using ModelTest-NG (32). The best-fit model for each alignment is as follows: LG+G4 for all multiple alignments of RT and RH proteins and LG+I+G4 for all multiple alignments of IN proteins.

Phylogeny congruence test. Jane 4 was used to compare the congruence between the phylogenies of Odin retrotransposons and their hosts (33). Three sets of cost values for cospeciation, duplication, duplication with host switch, loss, and failure to diverge were used, including 0, 1, 2, 1, 1; -1, 0, 0, 0, 0; and 0, 1, 1, 2, 0 (10). *P* values were estimated using the random parasite tree method with a sample size of 500.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, TXT file, 0.2 MB.
DATA SET S2, TXT file, 0.03 MB.
DATA SET S3, TXT file, 0.03 MB.
DATA SET S4, TXT file, 0.02 MB.
DATA SET S5, TXT file, 0.02 MB.
DATA SET S6, TXT file, 0.03 MB.
DATA SET S7, TXT file, 0.02 MB.
TABLE S1, PDF file, 0.1 MB.
TABLE S2, PDF file, 0.1 MB.
TABLE S3, PDF file, 0.1 MB.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (grant no. 31922001).

We declare no conflict of interest.

REFERENCES

1. Stoye JP. 2012. Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat Rev Microbiol 10:395–406. https://doi.org/10.1038/nrmicro2783.
2. Johnson WE. 2015. Endogenous retroviruses in the genomics era. Annu Rev Virol 2:135–159. https://doi.org/10.1146/annurev-virology-100114-054945.
3. Johnson WE. 2019. Origins and evolutionary consequences of ancient endogenous retroviruses. Nat Rev Microbiol 17:355–370. https://doi.org/10.1038/s41579-019-0189-2.
4. Gifford R, Tristem M. 2003. The evolution, distribution and diversity of endogenous retroviruses. Virus Genes 26:291–315. https://doi.org/10.1023/a:1024455415443.
5. Hayward A, Grabherr M, Jern P. 2013. Broad-scale phylogenomics provides insights into retrovirus-host evolution. Proc Natl Acad Sci U S A 110:20146–20151. https://doi.org/10.1073/pnas.1315419110.
6. Hayward A, Cornwallis CK, Jern P. 2015. Pan-vertebrate comparative genomics unmasks retrovirus macroevolution. Proc Natl Acad Sci U S A 112:464–469. https://doi.org/10.1073/pnas.1414980112.
7. Xu X, Zhao H, Gong Z, Han GZ. 2018. Endogenous retroviruses of non-avian/mammalian vertebrates illuminate diversity and deep history of retroviruses. PLoS Pathog 14:e1007072. https://doi.org/10.1371/journal.ppat.1007072.
8. Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res 46:D708–D717. https://doi.org/10.1093/nar/gkx932.
9. Gifford RJ, Blomberg J, Coffin JM, Fan H, Heidmann T, Mayer J, Stoye J, Tristem M, Johnson WE. 2018. Nomenclature for endogenous retrovirus (ERV) loci. Retrovirology 15:59. https://doi.org/10.1186/s12977-018-0442-1.
10. Wang J, Han GZ. 2021. A Sister lineage of sampled retroviruses corroborates the complex evolution of retroviruses. Mol Biol Evol 38:1031–1039. https://doi.org/10.1093/molbev/msaa272.
11. Malik HS, Eickbush TH. 2001. Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res 11:1187–1197. https://doi.org/10.1101/gr.185101.
12. Smyshlyaev G, Voigt F, Blinov A, Barabas O, Novikova O. 2013. Acquisition of an Archaea-like ribonuclease H domain by plant L1 retrotransposons supports modular evolution. Proc Natl Acad Sci U S A 110:20140–20145. https://doi.org/10.1073/pnas.1310958110.
13. Ustyantsev K, Novikova O, Blinov A, Smyshlyaev G. 2015. Convergent evolution of ribonuclease h in LTR retrotransposons and retroviruses. Mol Biol Evol 32:1197–1207. https://doi.org/10.1093/molbev/msv008.
14. Krupovic M, Blomberg J, Coffin JM, Dasgupta I, Fan H, Geering AD, Gifford R, Harrach B, Hull R, Johnson W, Kreuze JF, Lindemann D, Llorens C, Lockhart B, Mayer J, Muller E, Olszewski NE, Pappu HR, Pooggin MM, Richert-Poggeler KR, Sabanadzovic S, Sanfacon H, Schoelz JE, Seal S, Stavolone L, Stoye JP, Teycheney PY, Tristem M, Koonin EV, Kuhn JH. 2018. Ortervirales: new virus order unifying five families of reverse-transcribing viruses. J Virol 92:e00515-18. https://doi.org/10.1128/JVI.00515-18.
15. Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362. https://doi.org/10.1002/j.1460-2075.1990.tb07536.x.
16. Doolittle RF, Feng DF, Johnson MS, McClure MA. 1989. Origins and evolutionary relationships of retroviruses. Q Rev Biol 64:1–30. https://doi.org/10.1086/416128.
17. Eickbush TH, Jamburuthugoda VK. 2008. The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 134:221–234. https://doi.org/10.1016/j.virusres.2007.12.010.
18. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. 2007. A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982. https://doi.org/10.1038/nrg2165.
19. Dodonova SO, Prinz S, Bilanchone V, Sandmeyer S, Briggs JAG. 2019. Structure of the Ty3/Gypsy retrotransposon capsid and the evolution of retroviruses. Proc Natl Acad Sci U S A 116:10048–10057. https://doi.org/10.1073/pnas.1900931116.
20. Hayward A. 2017. Origin of the retroviruses: when, where, and how? Curr Opin Virol 25:23–27. https://doi.org/10.1016/j.coviro.2017.06.006.
21. Llorens C, Futami R, Covelli L, Dominguez-Escriba L, Viu JM, Tamarit D, Aguilar-Rodriguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A, Moya A. 2011. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res 39:D70–4. https://doi.org/10.1093/nar/gkq1061.
22. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010.
23. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
24. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res 45:D200–D203. https://doi.org/10.1093/nar/gkw1129.
25. Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35:W265–8. https://doi.org/10.1093/nar/gkm286.
26. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053.
27. Pei J, Kim BH, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucleic Acids Res 36:2295–2300. https://doi.org/10.1093/nar/gkn072.
28. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015.
29. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285.
30. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.
31. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542. https://doi.org/10.1093/sysbio/sys029.
32. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol Biol Evol 37:291–294. https://doi.org/10.1093/molbev/msz189.
33. Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. 2010. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms Mol Biol 5:16. https://doi.org/10.1186/1748-7188-5-16.
34. Kayal E, Roure B, Philippe H, Collins AG, Lavrov DV. 2013. Cnidarian phylogenetic relationships as revealed by mitogenomics. BMC Evol Biol 13:5. https://doi.org/10.1186/1471-2148-13-5.
35. Kayal E, Bentlage B, Pankey MS, Ohdera AH, Medina M, Plachetzki DC, Collins AG, Ryan JF. 2018. Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. Bmc Evol Biol 18:1–18. https://doi.org/10.1186/s12862-018-1142-0.
36. Pisani D, Pett W, Dohrmann M, Feuda R, Rota-Stabelli O, Philippe H, Lartillot N, Worheide G. 2015. Genomic data do not support comb jellies as the sister group to all other animals. Proc Natl Acad Sci U S A 112:15402–15407. https://doi.org/10.1073/pnas.1518127112.
37. Simakov O, Kawashima T, Marletaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, Yue JX, Chen YC, Freeman RM, Jr., Sasaki A, Hikosaka-Katayama T, Sato A, Fujie M, Baughman KW, Levine J, Gonzalez P, Cameron C, Fritzenwanker JH, Pani AM, Goto H, Kanda M, Arakaki N, Yamasaki S, Qu J, Cree A, Ding Y, Dinh HH, Dugan S, Holder M, Jhangiani SN, Kovar CL, Lee SL, Lewis LR, Morton D, Nazareth LV, Okwuonu G, Santibanez J, Chen R, Richards S, Muzny DM, Gillis A, Peshkin L, Wu M, Humphreys T, Su YH, Putnam NH, et al. 2015. Hemichordate genomes and deuterostome origins. Nature 527:459–465. https://doi.org/10.1038/nature16150.
38. Simion P, Philippe H, Baurain D, Jager M, Richter DJ, Di Franco A, Roure B, Satoh N, Queinnec E, Ereskovsky A, Lapebie P, Corre E, Delsuc F, King N, Worheide G, Manuel M. 2017. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. Curr Biol 27:958–967. https://doi.org/10.1016/j.cub.2017.02.031.
39. Gusmao LC, Van Deusen V, Daly M, Rodriguez E. 2020. Origin and evolution of the symbiosis between sea anemones (Cnidaria, Anthozoa, Actiniaria) and hermit crabs, with additional notes on anemone-gastropod associations. Mol Phylogenet Evol 148:106805. https://doi.org/10.1016/j.ympev.2020.106805.