Small circRNAs with self-cleaving ribozymes are frequently expressed in metazoan transcriptomes

Amelia Cervera, Marcos de la Peña*
IBMCP (CSIC-UPV). C/ Ingeniero Fausto Elio s/n, 46022, Valencia, Spain.

Marcos de la Peña (corresponding author*): rivero@ibmcp.upv.es

Running title: Metazoan circRNAs with ribozymes

Keywords: circular RNA, retrotransposons

Abbreviations: circRNA, circular RNA. HHR, hammerhead ribozyme. LINE, long interspersed element. LTR, long terminal repeat. PBS, primer binding site. PLE, Penelope-like element. PPT, polypurine tract. RT, retrotranscriptase. TSD, target site duplication.

Abstract

Ribozymes are catalytic RNAs present in the genomes of modern organisms but considered as remnants of a prebiotic RNA world. The paradigmatic Hammerhead ribozyme (HHR) is a small self-cleaving motif previously found widespread from bacterial to human genomes. Here, we report that classical type I HHRs belong to a novel family of non-autonomous non-LTR retrotransposons common in the genomes of diverse metazoan phyla, such as cnidarians, molluscs, arthropods or amphibians, among others. We demonstrate that these retroelements express as abundant and heterogeneous covalently closed circRNAs (~170-400 nt) in all the analysed tissues from diverse animal species. Our bioinformatic mining also revealed a novel class of type III HHRs in the genomes of diverse hymenopters and coleopters. These new self-cleaving motifs show large pseudoknot interactions and are associated to plant-like LTR retrozymes encoding larger circRNAs. Phylogenetic dating and the similarities between plant and animal LTR retrozymes suggest events of horizontal transfer from pollinating animals to plant genomes. In conclusion, our results reveal a new and conserved pathway through autocatalytic motifs for the biosynthesis of circRNAs, which naturally accumulate in the eukaryotic transcriptomes and show an intriguing capability of crossing interspecies barriers.

Introduction

The discovery of catalytic RNAs or ribozymes in the 80's [1, 2] strongly supported the hypothesis of a prebiotic RNA world, where the first living organisms were based on RNA as both the genetic material and as catalyst [3]. Some of these ancient ribozymes are thought to subsist in modern organisms, carrying out crucial biological roles such as the peptide bond formation [4], tRNA processing [2] or mRNA splicing [5]. There is, however, an enigmatic and less essential group of small (~50-150 nt) self-cleaving ribozymes with uncertain origins and functions [6]. The hammerhead ribozyme (HHR), the first and best-studied member of this family, has a conserved core of 15 nucleotides surrounded by three helixes (helix I to III), which adopt a γ-shaped fold where helix I interacts with helix II through characteristic tertiary interactions required for efficient self-cleavage [7-9]. There are three possible circularly permuted forms named type I, II or III depending on the open-ended helix of the motif (Figure 1A). The HHR was firstly discovered in infectious circRNAs of plants, such as viral RNA satellites and viroids [10], but also in the repetitive DNA of some amphibians [11], and, later on, in schistosomes [12] and cave crickets [13]. Few years ago, however, HHR motifs were found ubiquitously across all domains of life, from bacterial to eukaryotic genomes [14, 15], including humans [16], unveiling this catalytic RNA as the most widespread small ribozyme. The biological roles of these genomic HHRs, however, have remained poorly known. Rare minimal variants of the type I HHR were recently found encoded in retrotransposons of the Penelope (PLEs) [17] and Terminon [18] families, whereas a few highly conserved HHRs in mammals and other amniotes seem to play a role in mRNA biogenesis [19]. More recently, type III HHRs present in the genomes of several flowering plants were found involved in the transcript processing of a novel family of

non-autonomous LTR retrotransposons, so-called *retrozymes* (Figure 1B), which spread through circRNA intermediates of 600-800 nt [20].

In this work, canonical type I HHRs present in metazoan genomes have been thoroughly analyzed. Most of these genomic HHRs in animals can be found within short tandem repeats of a few hundred base pairs (150 to 450 bp) and following a patchy distribution among metazoan phyla [19]. Sequence repeats do not show any sequence identity between distant species, or clear protein-coding capacity. In contrast to LTR retrozymes from angiosperms, the metazoan repeats lack the characteristic LTRs, PBS and PPT motifs of plant retroelements, show lower GC content (below 50%), and, when present, the target side duplications (TSDs) are larger than 4 bp (Figure 1C). All these peculiarities indicate that metazoan repeats with HHRs are a new family of non-autonomous non-LTR retrozymes, which would be mobilized by autonomous non-LTR retrotransposons.

Small circRNAs with HHRs are expressed by non-LTR retrozymes from cnidarians to arthropods

Bioinformatic analysis of the genome of a stony coral such as *A. millepora* [21] show the presence of more than 6,000 *bona fide* type I motifs (either minimal or canonical. Figure 1A), which based on their primary sequence can be classified in several families (Figure S1A). Up to 1,508 of these motifs correspond to canonical type I HHRs, showing typical helix sizes and the precise tertiary interactions between loops 1 and 2 described for this ribozyme [9], whereas the other motifs resembled to the atypical minimal HHR variants already related to Penelope and Terminon classes of retrotranposons [17, 18]. The canonical HHRs of *A. millepora* are usually present within repeats of small size (220-240 bp), which show efficient self-cleavage during *in vitro*

transcription (Figure S1B). Northern blot analysis of coral RNA extracts under native conditions show the presence of a single band above 200 nt size, whereas the same extracts run under denaturing conditions show that these RNAs split into linear (~220-240 nt) and slow migrating bands (~300 nt), a characteristic behaviour of circular RNA molecules (Figure 2A). Enhanced separation of circular and linear molecules is obtained when ionic strength is lowered (from 1xTBE to 0.25xTBE) under the same denaturing conditions (Figure S1C). RT-PCR with divergent oligos confirmed the circular nature of PAGE-purified RNAs, which are predicted to adopt a highly self-paired secondary structure (Figure 2A). The genomes of stony corals from different families show the presence of retrozymes with similar sequence and secondary structure to the *A. millepora* ones (Figure S1D and E), indicating a phylogenetic relationship among all of them. Moreover, small repeats of ~230-350 bp carrying type I HHRs can be identified in the genomes of diverse jellyfishes and anemones, indicating that small circRNAs with HHRs are frequently encoded in the genomes of cnidarians (Figure S1F).

Type I HHRs within tandem repeats are also detected in the genomes of more complex metazoans, from molluscs to arthropods. The Mediterranean mussel *Mytilus galloprovincialis* contains arround 5,000 copies of a typical and fast cleaving type I HHR (Figure S2A) located within sequence repeats of 360-390 bp. Northern blot analysis of different mussel tissues show the presence of high levels of circRNAs corresponding to these repeats, most notably in gonads and mantle (Figure 2B). RT-PCR and cloning of purified circRNAs from the gonads of a single specimen show a population of quite heterogeneous sequences and sizes (indels) (Figure S2B). Most likely, the circRNAs in a given tissue could result from the expression of slightly different genomic copies. However, none of the obtained sequences exactly match with database entries corresponding to *M. galloprovincialis*, which suggests that, as

previously reported for plant LTR retrozymes [20], other explanations to this sequence heterogeneity can be possible (ie. RNA editing). As observed for the *A. millepora* genome, *M. galloprovincialis* also contain more than 3,000 motifs corresponding to minimal HHR variants typical of autonomous PLE and Terminon retroelements. Our bioinformatic searches confirm that analogous repeats with type I HHRs are encoded in many other invertebrate genomes, such as rotifers, platyhelminths, annelids, crustaceans or insects, which most likely express as small circRNAs (~170-400 nt) adopting similar highly self-paired secondary structures (Figure 3).

The axolotl salamander contain up to 126,000 retrozyme copies encoding two different families of circRNAs

The first metazoan type I HHR was already described in 1987 in the so-called *satellite DNA* repeats (~330 bp) of newts and salamanders [11]. Due to a short helix III, these ribozymes showed low catalytic activity as single motifs, which suggested that they may self-cleave *in vivo* as dimers [22]. The satellite DNA of these amphibians was reported to express in diverse tissues, most notably in gonads, as monomeric and multimeric linear RNAs, which carryed 5'-hydroxyls and phosphate blocked 3' ends [23]. Our analyses of the recently sequenced genome of the axolotl salamander (*Ambystoma mexicanum*) [24] reveal the presence of 126,247 motifs of the canonical type I HHR. These ribozymes correspond to two different HHR sequences, each one associated with a slightly different repeat of 330 and 350 bp, respectively (Figure S3). The axolotl ribozymes show very weak self-processing activity as monomeric motifs, whereas the dimeric constructs results in much more efficient self-cleavage (Figure S3C). As previously described for other salamanders [23], northern blots of axolotl RNA extracts performed under native conditions confirm the presence of monomeric

and multimeric RNAs of the size encompassed by the HHRs (Figure 2C). When the same extracts were run under denaturing conditions, however, we confirm that RNA populations are composed of a mixture of linear and, most notably, circular molecules of 330 and 350 nt (Figure 2C). *In vitro* transcription of dimeric retrozyme constructs from *A. mexicanum* allowed us to purify monomeric linear RNAs after double self-cleavage. These RNAs were assayed for either self- or tRNA ligase-circularization. Although self-circularization can be readily detected, final levels are very low (3%) compared with the ones obtained after incubation with the RtcB tRNA ligase (>80%) (Figure S3D). These observations indicate that, as proposed for plant retrozymes [20], a tRNA ligase would be responsible of retrozyme circularization in metazoans, in spite of the known capabilities of the *in vitro* self-ligation of the HHR [25]. The axolotl also contains numerous PLE retroelements [24] carrying minimal HHR copies in dimeric tandems similar to those found in retrozymes (Figure S3A and B), which suggests again that these particular autonomous retroelements would be involved in the transposition of non-LTR retrozymes.

Novel type III HHRs in plant-like retrozymes of pollinating insects

In a latter bioinformatic search, we thoroughly analysed the intriguing lack of type I and type III HHRs in the genomes of either plants or animals, respectively. We found out that whereas only a few examples of type I HHRs can be detected in plants [14, 17], numerous rare variants of type III HHRs can be detected within short tandem repeats in metazoans, suggesting new families of retrozymes. Freshwater hydras, such as *H. vulgaris* or *H. magnipapillata*, show hundreds of type III HHRs, which in many instances occur within non-LTR-like retrozymes composed of AU-rich sequences (~400 bp). These ribozymes show a larger helix I (up to 32 nt) with an internal loop that would

mimic the tertiary interactions present in type I HHRs (Data not shown). More interestingly, highly conserved motifs of a novel class of type III HHRs were also detected in diverse hymenopters (ants, wasps, bees and bumblebees) and coleopters (beetle and ladybug). These new ribozymes preserve the the core sequences and stem sizes of canonical HHRs, but much larger loops 1 and 2 (13-23 nt instead ot the typical 4-8 nt), which allow extensive pseudoknot interactions between them (here after, pknot-HHRs) (Figure 4). The number of base pairs detected for these pseudoknots ranged from 10 to 13 Watson-Crick bp (including some G·U wobble pairs). This large loop-loop interaction totally differs from the one described for classical type III HHRs from subviral and plant genomes and composed by a few non-canonical interactions [26]. *In vitro* analysis of pknot-HHRs, however, reveal very low self-cleavage rates for these novel ribozymes, with observed rate constants more than one order of magnitude below the observed ones for classical type III HHRs under similar post- and co-transcriptional conditions (Figure 4A, Figure S4A) [7, 27]. Analyses of pknot-HHRs with either long or short helixes III result in similarly low cleavage rates, indicating that the equilibrium is not displaced to ligation, and that these pknot-HHRs under *in vitro* conditions are intrinsically slow self-cleaving motifs. Mutagenesis of the loop sequences keeping the number of W-C interactions results in a similar self-cleavage efficiency, which indicates that sequence conservation of the pseudoknots may have other role different from cleavage (Figure S4). Interestingly, genomic pknot-HHRs can occur as dimers within small LTRs (~250bp) flanked by 4 bp TSDs, a PBS ($tRNA^{Trp}$) and a PPT motif, plus a variable non-coding region of 500-800 bp (>55% GC) between ribozymes (Figure 3C). These features match with those of plant retrozymes (Figure 1B), which indicates that metazoan LTR retrozymes with pknot-HHRs will propagate through similar circRNA intermediates of 500-800 nt (Figure S4E) to be mobilized by the machinery of

autonomous LTR retrotransposons of the insect. Moreover, sequence alignments of the pknot-HHRs reveals a clear evolutionary conservation of the whole ribozyme motif among most insects (Figure 4D), a feature that was not observed in the HHRs of plant retrozymes [20]. This phylogenetic relationship between pknot-HHRs suggests that these motifs, and their associated LTR retrozymes, should have already existed in the ancestors of coleopters and hymenopters, which diverged ~350 Mya [28]. This date is far before the estimated origin of plant dicots (~120 Mya) or even flowering plants (~140-200 Mya) [29], allowing us to propose that plants would have acquired their LTR retrozymes by horizontal transfer from pollinating insects. Although other explanations are also possible (ie. convergent evolution), the patchy distribution observed for plant retrozymes [20], and the precise geographic localizations of retrozyme-containing species (most of the insects and plants found to carry potentially active retrozymes are located in the American continent) support the assumption of one or more events of horizontal transfer of this retroelement through circRNAs as the most likely vector.

Discussion

The results reported in this work answer a longstanding question about the presence of a paradigmatic small ribozyme such as the HHR, in animal genomes. The first HHRs in a metazoan genome were reported 32 years ago in newts [11], and afterwards, also in trematodes [12] and cave crickets [13]. Although further characterization of amphibian repeats indicated that they expressed as linear transcripts [23], now we know that a larger fraction of the purified RNA is composed by circular molecules. Moreover, we found that this particular phenomenon is taking place in animals from different phyla, which combined with our bioinformatic analysis, allow us to conclude that DNA repeats with HHRs in metazoans and eukaryotes in general, are a peculiar mobile element that

express abundantly as heterogeneous circular RNAs in all somatic and germinal tissues analyzed. A previous model for the life cycle of plant retrozymes [20] could be extended for all eukaryotic retrozymes, which would be initially transcribed and HHR self-processed, following by covalent circularization by host RNA ligases. The resulting circRNAs can be regarded as intermediates of retrotransposition and, among other uncertain roles in the cell for such abundant RNAs or ribonucleoprotein complexes [30], they would act as templates for retrotranscriptases encoded by autonomous LTR (ie. Gypsy/Copia for type III HHR-containing retrozymes) or non-LTR (ie. PLEs for type I HHR-containing retrozymes) retroelements. Due to the circular nature of retrozyme RNAs, their retrotranscription results in multimeric cDNAs, which are integrated in new genomic loci by the retrotransposon machinery. In summary, our work reveals the existence of a new pathway for circRNA biosynthesis in eukaryotes through autocatalytic RNAs, in addition to the previously reported circRNA pathway through backsplicing [31]. Future research on RNA circles with autocatalytic ribozymes will help us to decipher the roles of these abundant molecules in eukaryotes, but also to develop new biotechnological tools in the emerging field of circRNAs.

Methods

Bioinformatics

RNAmotif software was used for the detection of either canonical, minimal or pseudoknotted HHR motifs (either type-I or type-III architectures) in metazoan genomes (RefSeq Genome Database, invertebrate/ and vertebrate_other/, plus some other individual genomes based on the obtained results). Specific descriptors (giving variable sizes of loops and stems) were defined for each HHR motif searches. The hits obtained were inspected for either the presence of tertiary interactions between helix I

and II to ensure they were *bona fide* HHRs or their occurrence within tandem repeats. Sequence homology searches through BLAST and BLASTX tools were carried out in individual or grouped genomes. Sequence alignments were performed with ClustalX and Jalview software. Secondary RNA structures of minimum free energy were calculated with the RNAfold program from the ViennaRNA Package and depicted with RNAviz.

Nucleic acid extractions

DNA from animal tissues was extracted following either a CTAB-based protocol with some modifications [20]. Briefly, samples were homogenised (1:2.5 w/v) in extraction buffer (1% w/v CTAB, 50 mM Tris-HCl pH 8, 10 mM EDTA, 0.7 M NaCl, 1% w/v PVP-40, 1% 2-mercaptoethanol), and were incubated for 30 min at 60ºC. An equal volume of chloroform:isoamyl alcohol (24:1 v/v) was added to the samples, which were then centrifuged. DNA in the supernatant was precipitated with 0.1 volumes of 3 M sodium acetate (pH 5.5) and 2.5 volumes of 100% ethanol, dissolved in MilliQ water (Millipore), and quantified in a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

RNA from animal tissues was isolated with Trizol reagent (Invitrogen), following the manufacturer's instructions. Shortly, tissues were homogenised (1:10 v/v) in Trizol, and then 0.2 volumes of chloroform:isoamyl alchohol (24:1 v/v) were added to the samples. The mixture was centrifuged, and RNA in the supernatant was precipitated with 0.1 volumes of 3 M sodium acetate (pH 5.5) and 2.5 volumes of 100% ethanol. RNA was resuspended and quantified as explained above.

PCR, RT-PCR and molecular cloning

Genomic retrozymes were amplified by PCR using degenerate primers designed to bind the most conserved motifs within a species. The hot-start, proof-reading PrimeStar HS DNA Polymerase (Takara) was used following the manufacturer's instructions, together with phosphorylated (T4 polynucleotide kinase, Takara) primers that spanned the length of the RNA of the retrozyme (i.e., the retrozyme monomer). Amplification products of the adequate size were extracted from native 5% 1× TAE polyacrylamide gel slices with phenol:chloroform:isoamyl alcohol (25:24:1), and concentrated by ethanol precipitation as described above.

For circular retrozyme RNA cloning by RT-PCR, RNA was fractionated by native 5% 1× TAE PAGE, and extracted from gel slices corresponding to the retrozyme size (300 – 500 nt for *M. galloprovincialis*). RNAs were reverse-transcribed with SuperScript II (Invitrogen), and PCR-amplified with adjacent divergent primers previously phosphorylated, using Prime Star HS DNA polymerase.

PCR products were cloned into a linearised pBlueScript KS (Agilent) vector by blunt-end ligation, and were sequenced automatically with an ABI Prism DNA sequencer (Perkin–Elmer).

RNA transcriptions

RNAs of the retrozyme sequences were synthesized by *in vitro* transcription of pBlueScript KS plasmids containing the corresponding retrozyme fragment or HHR motif linearized with the appropriate restriction enzyme. Transcription reactions contained: 40 mM Tris–HCl, pH 8, 6 mM MgCl2, 2 mM spermidine, 0.5 mg/ml RNase-free bovine serum albumin, 0.1% Triton X-100, 10 mM dithiothreitol, 1 mM each of CTP, GTP and UTP, 0.1 mM ATP plus 0.5 µCi/µl [α-32P]ATP, 2 U/µl of human placental ribonuclease inhibitor, 20 ng/µl of plasmid DNA, and 4 U/µl of T7 or T3

RNA polymerases. After incubation at 37 °C for 2 h, products were fractionated by polyacrylamide gel electrophoresis (PAGE) in 5% (retrozymes) or 15% gels (ribozymes) with 8 M urea.

Kinetic analysis of self-cleavage under post- and co-transcriptional conditions

For determining the cleaving rate constants, uncleaved primary transcripts of pknot-HHRs (~1 nM) were incubated in 20 µL of 50 mM Tris-HCl (pH 8) for 1 min at 95°C and slowly cooled for 30 min to 25°C. After taking a zero-time aliquot, self-cleavage reactions were triggered by adding $MgCl_2$ to 10 mM (initial assays done at physiological Mg concentrations of 1 mM resulted in very slow kinetics of cleavage). Aliquots were removed at appropriate time intervals and quenched with a fivefold excess of stop solution at 0°C. The substrates and cleavage products were separated by PAGE in 15% denaturing gels. The product fraction at different times, Ft, was determined by quantitative scanning of the corresponding gel bands and fitted to the equation $F_t = F_o + F_\infty(1 - e^{-kt})$, where F_o and F_∞ are the product fractions at zero time and at the reaction endpoint, respectively, and k is the first-order rate constant of cleavage (k_{obs}). Due to the very low kinetics of cleavage observed for pknot-HHRs, we deduced that unfolding/folding issues of these motifs could be affecting the observed results. Thus, we performed analyses under co-transcriptional conditions as previously described [27]. Briefly, transcription reactions were carried out at 37°C as described before, and appropriate aliquots were removed at different time intervals, quenched and analyzed as previously described [27].

Northern blot analysis

For northern blot analysis, from 5 up to 50 μg of purified RNA from the different animal tissues were examined in 5 % polyacrylamide gels under native (1×TAE) or denaturing (8 M urea, at either 0.25×TBE or 1×TBE) conditions. After ethidium bromide staining, RNAs were electroblotted to nylon membranes (Amersham Hybond-N, GE Healthcare) and UV-fixed with a crosslinker (UVC 500, Hoefer). Prehybridization, hybridization (at 68°C in 50% formamide) and washing was done following the instructions of the manufacturer (GE Healthcare). The riboprobes were obtained by transcription of linearized pBlueScript plasmid containing the full retrozyme in the presence of DIG-UTP (Roche Diagnostic GmbH) [20].

Acknowledgments

We acknowledge the excellent technical assistance of M. Pedrote. Funding for this work was provided by the Ministerio de Ciencia, Innovación y Universidades of Spain and FEDER funds (BFU2017-87370-P).

Author contributions

A.C. and M.P. designed and performed the experiments; M.P. wrote the manuscript; A.C. and M.P. interpreted the data and reviewed the text and figures.

References

1. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell *31*, 147-157.
2. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell *35*, 849-857.
3. Gilbert, W. (1986). The RNA World. Nature *319*, 618.

4. Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. Science *289*, 920-930.
5. Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. Nature *413*, 701-707.
6. Jimenez, R.M., Polanco, J.A., and Luptak, A. (2015). Chemistry and Biology of Self-Cleaving Ribozymes. Trends Biochem Sci *40*, 648-661.
7. De la Peña, M., Gago, S., and Flores, R. (2003). Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. EMBO J *22*, 5561-5570.
8. Khvorova, A., Lescoute, A., Westhof, E., and Jayasena, S.D. (2003). Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol *10*, 708-712.
9. Martick, M., and Scott, W.G. (2006). Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell *126*, 309-320.
10. Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R., and Bruening, G. (1986). Autolytic processing of dimeric plant virus satellite RNA. Science *231*, 1577-1580.
11. Epstein, L.M., and Gall, J.G. (1987). Self-cleaving transcripts of satellite DNA from the newt. Cell *48*, 535-543.
12. Ferbeyre, G., Smith, J.M., and Cedergren, R. (1998). Schistosome satellite DNA encodes active hammerhead ribozymes. Mol Cell Biol *18*, 3880-3888.
13. Rojas, A.A., Vazquez-Tello, A., Ferbeyre, G., Venanzetti, F., Bachmann, L., Paquin, B., Sbordoni, V., and Cedergren, R. (2000). Hammerhead-mediated processing of satellite pDo500 family transcripts from Dolichopoda cave crickets. Nucleic Acids Res *28*, 4037-4043.
14. De la Peña, M., and Garcia-Robles, I. (2010). Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. RNA *16*, 1943-1950.
15. Perreault, J., Weinberg, Z., Roth, A., Popescu, O., Chartrand, P., Ferbeyre, G., and Breaker, R.R. (2011). Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. PLoS computational biology *7*, e1002031.
16. De la Peña, M., and Garcia-Robles, I. (2010). Intronic hammerhead ribozymes are ultraconserved in the human genome. EMBO Rep *11*, 711-716.
17. Cervera, A., and De la Peña, M. (2014). Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. Molecular biology and evolution *31*, 2941-2947.
18. Arkhipova, I.R., Yushenova, I.A., and Rodriguez, F. (2017). Giant Reverse Transcriptase-Encoding Transposable Elements at Telomeres. Molecular biology and evolution *34*, 2245-2257.
19. De la Peña, M., Garcia-Robles, I., and Cervera, A. (2017). The Hammerhead Ribozyme: A Long History for a Short RNA. Molecules *22*.
20. Cervera, A., Urbina, D., and de la Peña, M. (2016). Retrozymes are a unique family of non-autonomous retrotransposons with hammerhead ribozymes that propagate in plants through circular RNAs. Genome Biol *17*, 135.
21. Fuller, Z.L.L., Y.; Bay, L.; Matz, M.; Przeworski M. (2018). Acropora millepora Genome Assembly. https://przeworskilab.com/wp-content/uploads/acropora-millepora-v2.pdf (to be submitted).
22. Forster, A.C., Davies, C., Sheldon, C.C., Jeffries, A.C., and Symons, R.H. (1988). Self-cleaving viroid and newt RNAs may only be active as dimers. Nature *334*, 265-267.
23. Epstein, L.M., and Coats, S.R. (1991). Tissue-specific permutations of self-cleaving newt satellite-2 transcripts. Gene *107*, 213-218.
24. Nowoshilow, S., Schloissnig, S., Fei, J.F., Dahl, A., Pang, A.W.C., Pippel, M., Winkler, S., Hastie, A.R., Young, G., Roscito, J.G., et al. (2018). The axolotl genome and the evolution of key tissue formation regulators. Nature *554*, 50-55.

25. Canny, M.D., Jucker, F.M., and Pardi, A. (2007). Efficient ligation of the Schistosoma hammerhead ribozyme. Biochemistry *46*, 3826-3834.
26. Chi, Y.I., Martick, M., Lares, M., Kim, R., Scott, W.G., and Kim, S.H. (2008). Capturing hammerhead ribozyme structures in action by modulating general base catalysis. PLoS Biol *6*, e234.
27. Long, D.M., and Uhlenbeck, O.C. (1994). Kinetic characterization of intramolecular and intermolecular hammerhead RNAs with stem II deletions. Proc Natl Acad Sci U S A *91*, 6977-6981.
28. Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B., Ware, J., Flouri, T., Beutel, R.G., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. Science *346*, 763-767.
29. Sauquet, H., von Balthazar, M., Magallon, S., Doyle, J.A., Endress, P.K., Bailes, E.J., Barroso de Morais, E., Bull-Herenu, K., Carrive, L., Chartier, M., et al. (2017). The ancestral flower of angiosperms and its early diversification. Nature communications *8*, 16047.
30. Luzi, E., Eckstein, F., and Barsacchi, G. (1997). The newt ribozyme is part of a riboprotein complex. Proc Natl Acad Sci U S A *94*, 9711-9716.
31. Li, X., Yang, L., and Chen, L.L. (2018). The Biogenesis, Functions, and Challenges of Circular RNAs. Mol Cell *71*, 428-442.

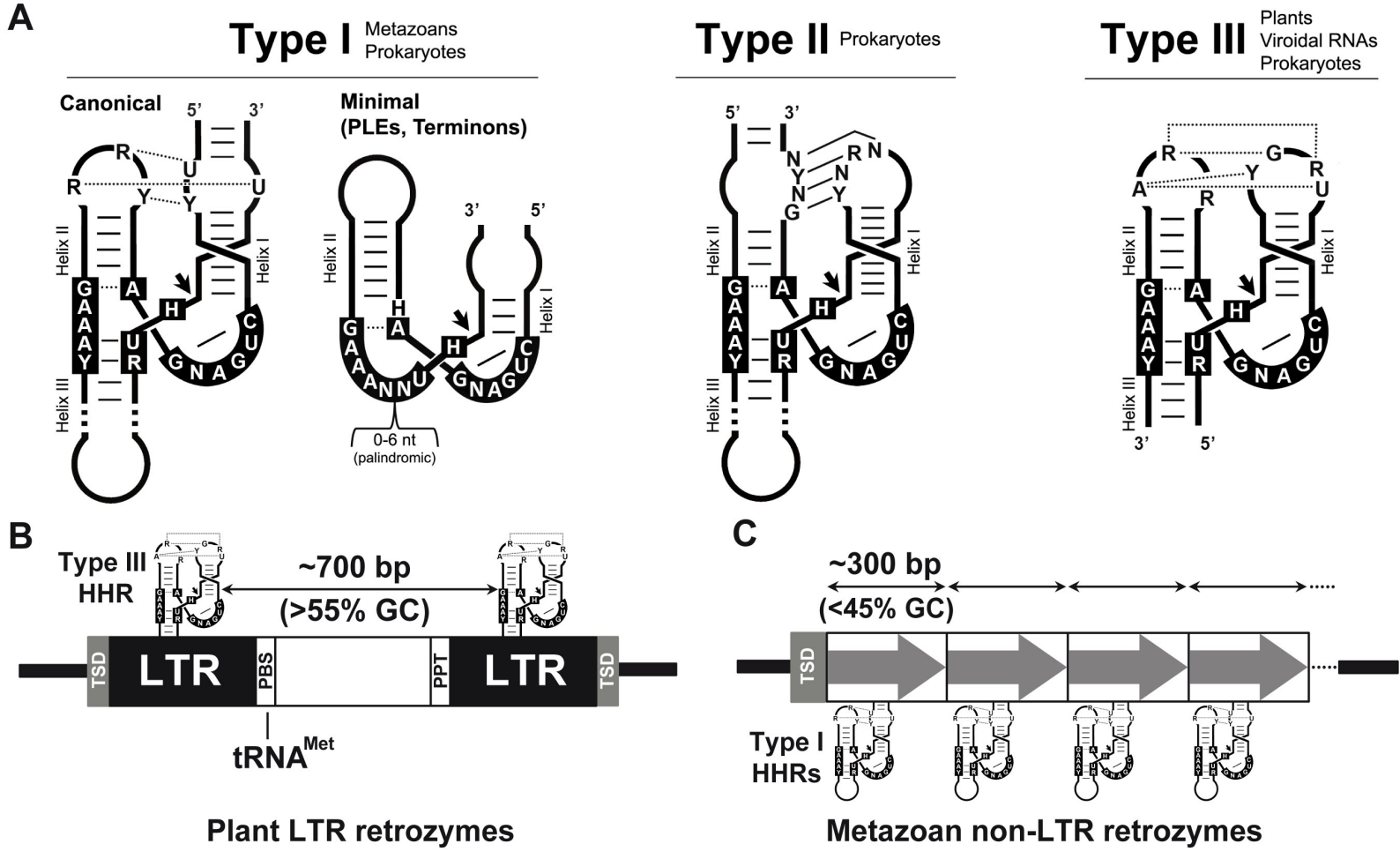

A
Type I
Metazoans
Prokaryotes
Type II
Prokaryotes
Type III
Plants
Viroidal RNAs
Prokaryotes
Canonical
Minimal
(PLEs, Terminons)
5'
3'
Helix I
Helix II
Helix III
0-6 nt
(palindromic)
B
Type III
HHR
~700 bp
(>55% GC)
TSD
LTR
PBS
PPT
LTR
TSD
tRNA^Met
Plant LTR retrozymes
C
~300 bp
(<45% GC)
TSD
Type I
HHRs
Metazoan non-LTR retrozymes

Figure 1.

(A) Schematic representation of the three possible hammerhead ribozyme (HHR) topologies (Type I, II and III). The most frequent nucleotides in the catalytic core (boxed) and in the loop-loop interactions are indicated. Dotted and continuous lines refer to Non-canonical and Watson-Crick base pairs. The three HHR types have been reported in Prokaryotic/Phages genomes, whereas metazoan and plant genomes mostly show type I and III motifs, respectively. (B) Schematic representation of a typical LTR retrozyme from plant genomes containing type III HHRs. The average size encompassed by the HHRs (~700 bp) and GC content (>55%) are indicated (C) Canonical type I HHRs in metazoan genomes can be found within short tandem repeats (~300 bp, from dimers to large multimers) of low GC content (<45%), which can be regarded as a new family of non-LTR retrozymes.

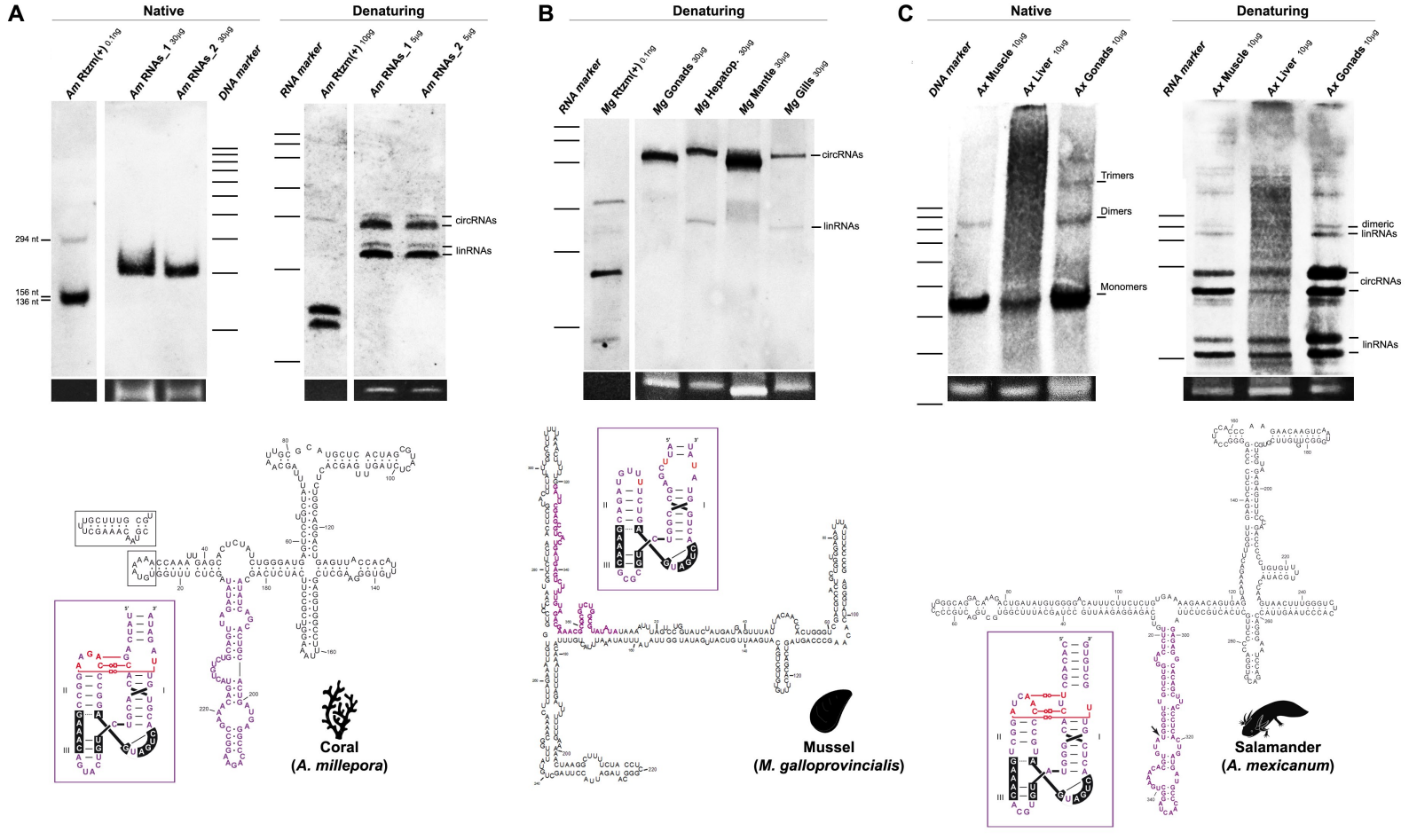

A
Native
Denaturing
Am Rtzm(+) 0.1ng
Am RNAs_1 30µg
Am RNAs_2 30µg
DNA marker
RNA marker
Am Rtzm(+) 10ng
Am RNAs_1 5µg
Am RNAs_2 5µg
294 nt
156 nt
136 nt
circRNAs
linRNAs
B
Denaturing
RNA marker
Mg Rtzm(+) 0.1ng
Mg Gonads 30µg
Mg Hepatop. 30µg
Mg Mantle 30µg
Mg Gills 30µg
circRNAs
linRNAs
C
Native
Denaturing
DNA marker
Ax Muscle 10µg
Ax Liver 10µg
Ax Gonads 10µg
Trimers
Dimers
Monomers
RNA marker
Ax Muscle 10µg
Ax Liver 10µg
Ax Gonads 10µg
dimeric linRNAs
circRNAs
linRNAs
Coral
(A. millepora)
Mussel
(M. galloprovincialis)
Salamander
(A. mexicanum)

Figure 2.

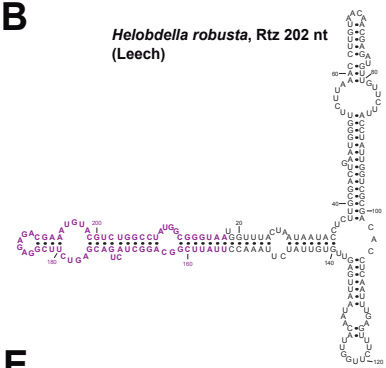
(A) Northern blot analysis of RNA extracts from two different samples of *Acropora millepora* coral that were run in the presence of a transcribed *A. millepora* retrozyme fragment under native (left membrane) and denaturing conditions (right membrane). On the right, minimum free energy secondary structure prediction of the circRNA derived from a representative retrozyme of the *A. millepora* genome. A typical 20 nt insertion found in some genomic retrozymes is indicated. The sequence of the HHR motif and its secondary structure is shown in purple letters. (B) Northern blot under denaturing conditions of RNA extracts from different tissues of the *Mytilus galloprovincialis* mussel (left) and the corresponding secondary structure of a cloned retrozyme circRNA from mussel gonads. (C) Northern blot analysis of RNA extracts from different tissues of the *Ambystoma mexicanum* salamander that were run under native (left membrane) and denaturing conditions (right membrane). On the right side, minimum free energy secondary structure prediction of a 350 nt circRNA derived from a retrozyme of the *A. mexicanum* genome. The approximate positions of the bands corresponding to the DNA 100-1,000bp (native blots) and RNA LR Riboruler (denaturing blots) markers, and the RNAs corresponding to circular and linear forms are indicated. Ethidium bromide staining of the 5S rRNAs are shown at the bottom as loading controls.

A

***Philodina roseola*, Rtz 174 nt**
(Rotifer)

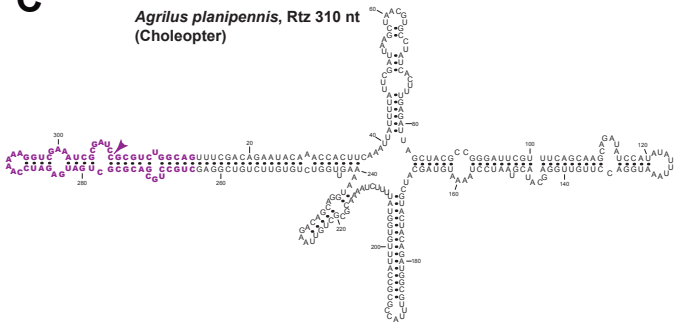
B

***Helobdella robusta*, Rtz 202 nt**
(Leech)



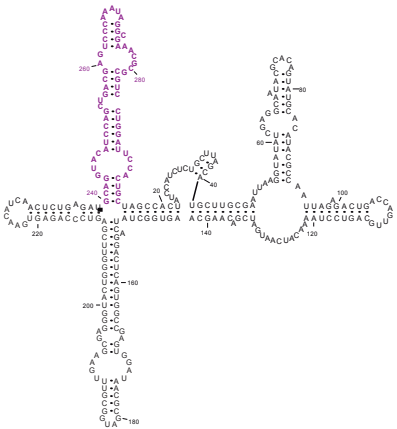
C

***Agrilus planipennis*, Rtz 310 nt**
(Choleopter)



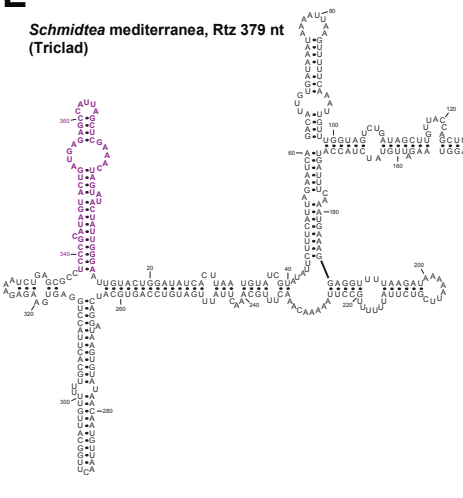
D

***Schistosoma mansoni*, Rtz 285 nt**
(Trematode)



E

***Schmidtea* mediterranea, Rtz 379 nt**
(Triclad)



F

***Eurytemora affinis*, Rtz 278 nt**
(Crustacean)

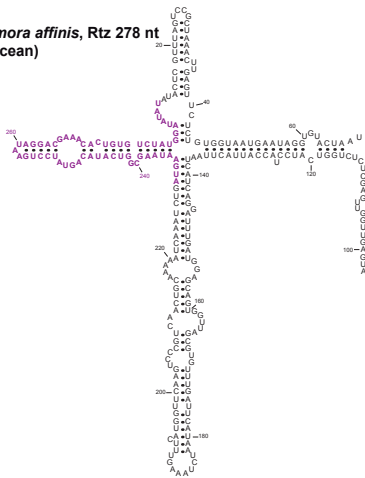


Figure 3

Secondary structure of minimum free energy of the representative circRNAs predicted to be encoded by retrozymes present in the genomes of invertebrates from different phyla across the eukaryotic tree of life, such as (A) rotifers, (B) trematodes, (C) insects, (D) annelids, (E) planarians or (F) crustaceans.

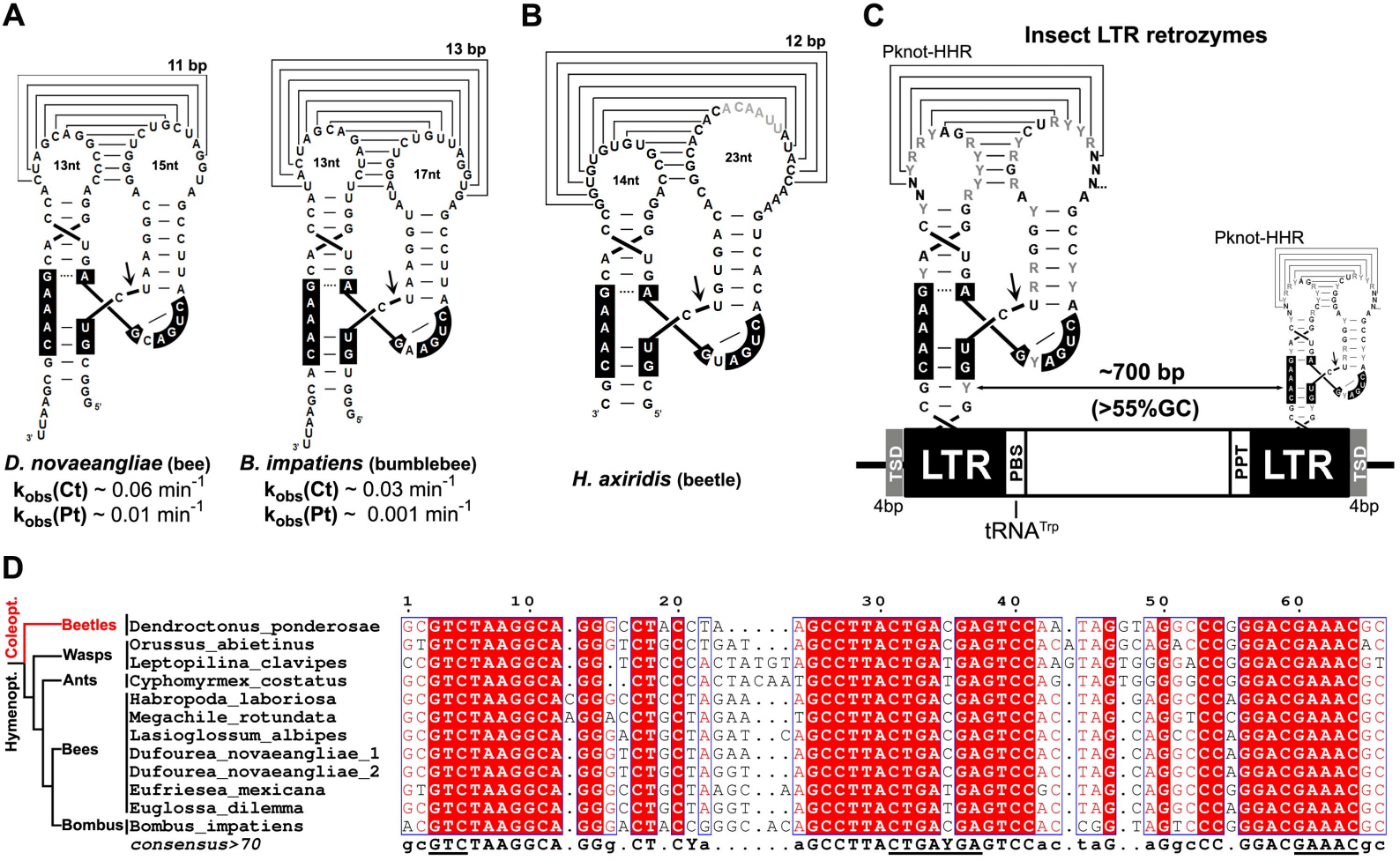
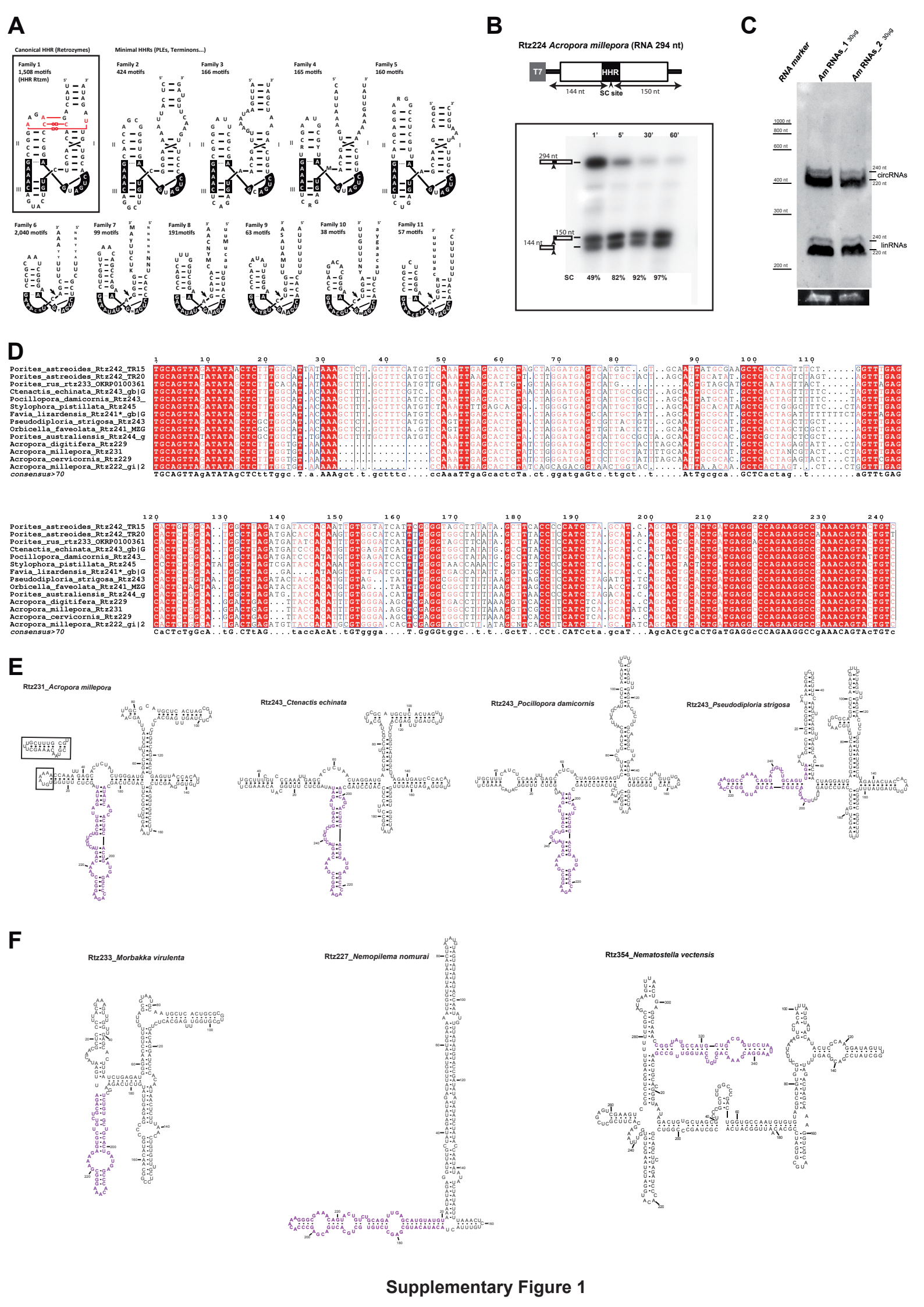

A
11 bp
13nt
15nt
D. novaeangliae (bee)
kobs(Ct) ~ 0.06 min-1
kobs(Pt) ~ 0.01 min-1
13 bp
13nt
17nt
B. impatiens (bumblebee)
kobs(Ct) ~ 0.03 min-1
kobs(Pt) ~ 0.001 min-1
B
12 bp
14nt
23nt
H. axiridis (beetle)
C
Insect LTR retrozymes
Pknot-HHR
Pknot-HHR
~700 bp
(>55%GC)
TSD
LTR
PBS
PPT
LTR
TSD
4bp
4bp
tRNATrp
D
Coleopt.
Hymenopt.
Beetles
Wasps
Ants
Bees
Bombus
Dendroctonus_ponderosae
Orussus_abietinus
Leptopilina_clavipes
Cyphomyrmex_costatus
Habropoda_laboriosa
Megachile_rotundata
Lasioglossum_albipes
Dufourea_novaeangliae_1
Dufourea_novaeangliae_2
Eufriesea_mexicana
Euglossa_dilemma
Bombus_impatiens
consensus>70

Figure 4.

(A) Two selected examples of type III hammerhead ribozymes with pseudoknots (pknot-HHRs) detected in hymenopteran genomes. *Dufourea novaeangliae* bee (left) and *Bombus impatiens* bumblebee (right) show pseudoknot interactions of 11 and 13 base pairs, respectively. The observed rates of cleavage for each motif under co-transcriptional (Ct) and post-transcriptional (Pt) conditions are shown in the bottom (see Methods for details). (B) A pknot-HHR in the genome of *Harmonia axyridis* ladybug is different to those found in most insects. Loop 1 shows a 6-8 nt insertion, which allows a pseudoknot interaction of 12 or, alternatively, 10 bp (see Figure S5E). (D) Schematic representation of the genomic LTR retrozymes with pknot-HHRs. A consensus sequence of the most frequent riboyzmes is shown. Average size and GC content of the sequence encompassed by the HHRs are indicated. (D) Sequence alignment of the pknot-HHRs from diverse hymenopters and the coleopter *Dendroctonus ponderosae*.



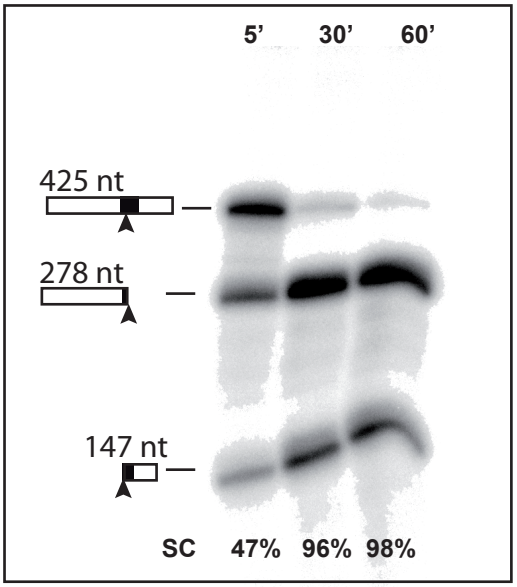
Supplementary Figure 1

Supplementary Figure 1

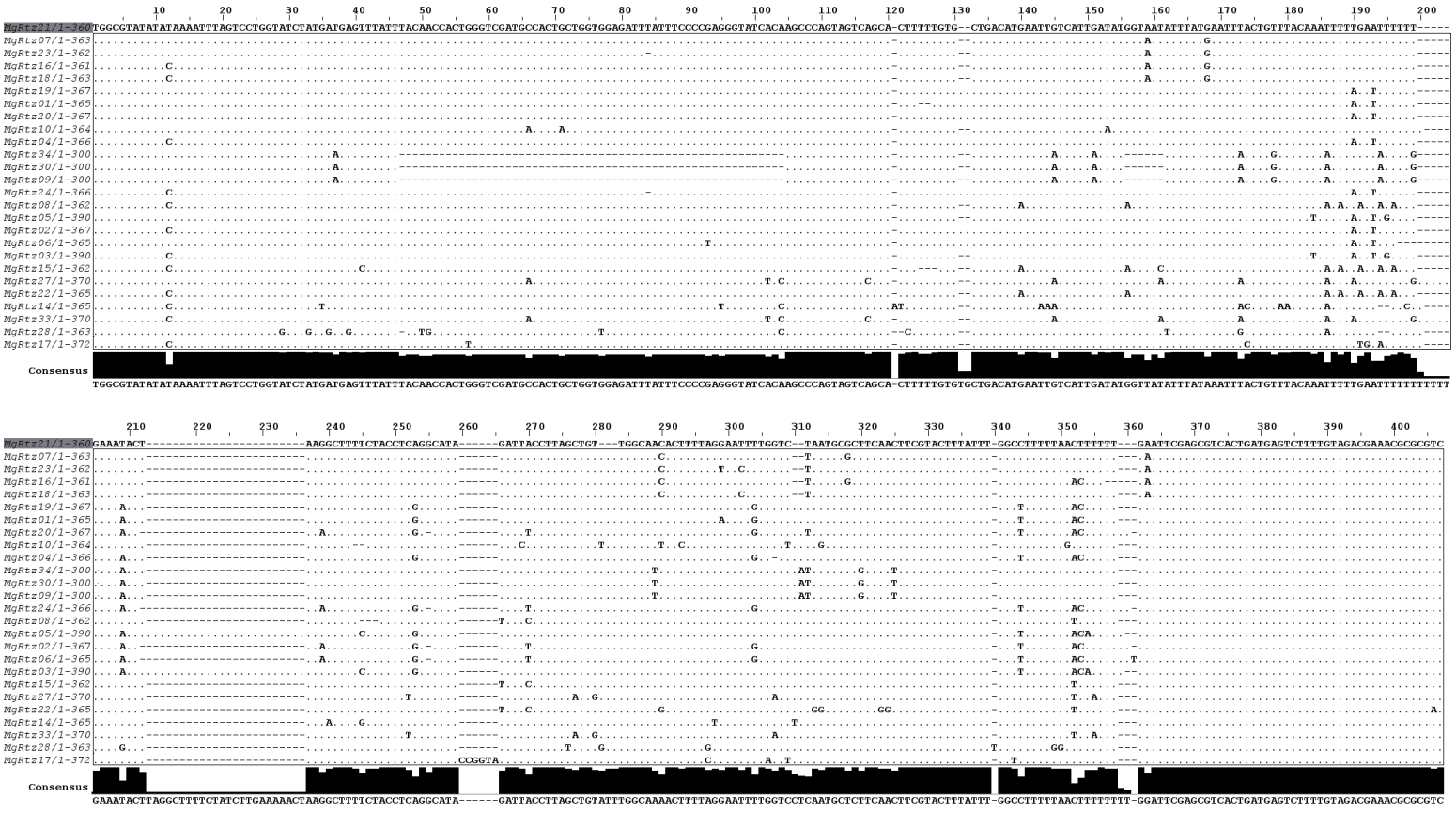
(A) Selection of the major type I HHR families detected in the genome of the *A. millepora* coral. Motifs of family 1 (boxed inset) correspond to the canonical type I HHR present in the non-autonomous retrozymes, whereas all the other minimal HHR motifs are related to autonomous retrotransposons, such as PLEs or Terminons. The consensus sequence (70%) of the aligned HHRs for each family is shown. Only those sequences allowing perfect helix topology and the conserved catalytic center (CUGANGA / GAAAY / RUH) were used for the HHR alignments. Nucleotides in red correspond with conserved loop-loop tertiary interactions of canonical type I HHRs. (B) A genomic copy of *A. millepora* retrozyme was cloned in a pBlueScript vector. Schematic representation of the clone and the expected sizes of the fragments after HHR self-cleavage are indicated. In the lower panel is shown an autoradiography of a run-off transcription at different times in the presence of [α-32P]-ATP. Cleavage of the whole RNA of the retrozyme reaches almost completion (95%). (C) Northern blot analysis of RNA extracts from two different samples of *Acropora millepora* corals that were run under denaturing conditions and low ionic strength (0.25xTBE). Bands corresponding to linear and circular RNAs are indicated. Approximated positions of RNAs from LR Riboruler RNA marker are indicated at the left. Ethidium bromide staining of the 5S rRNAs are shown at the bottom as loading controls. (D) Sequence alignment of selected retrozyme elements from 12 species of stony corals. The first positions of the sequences correspond to the self-cleavage site of the retrozymes. Highest sequence homology occurs among either the ribozyme motifs or the central region of the retrozymes (positions ~100-150 nt). (E) Minimum free energy secondary structure of some genomic retrozyme sequences from (E) selected stony corals and (F) other cnidarians such as jellyfishes (*Morbakka virulenta* and Nemopilema nomurai) and an anemone (*Nematostella vectensis).* Sequences corresponding to the hammerhead ribozymes are shown in purple.

A

Rtz375 *Mytilus galloprovincialis*



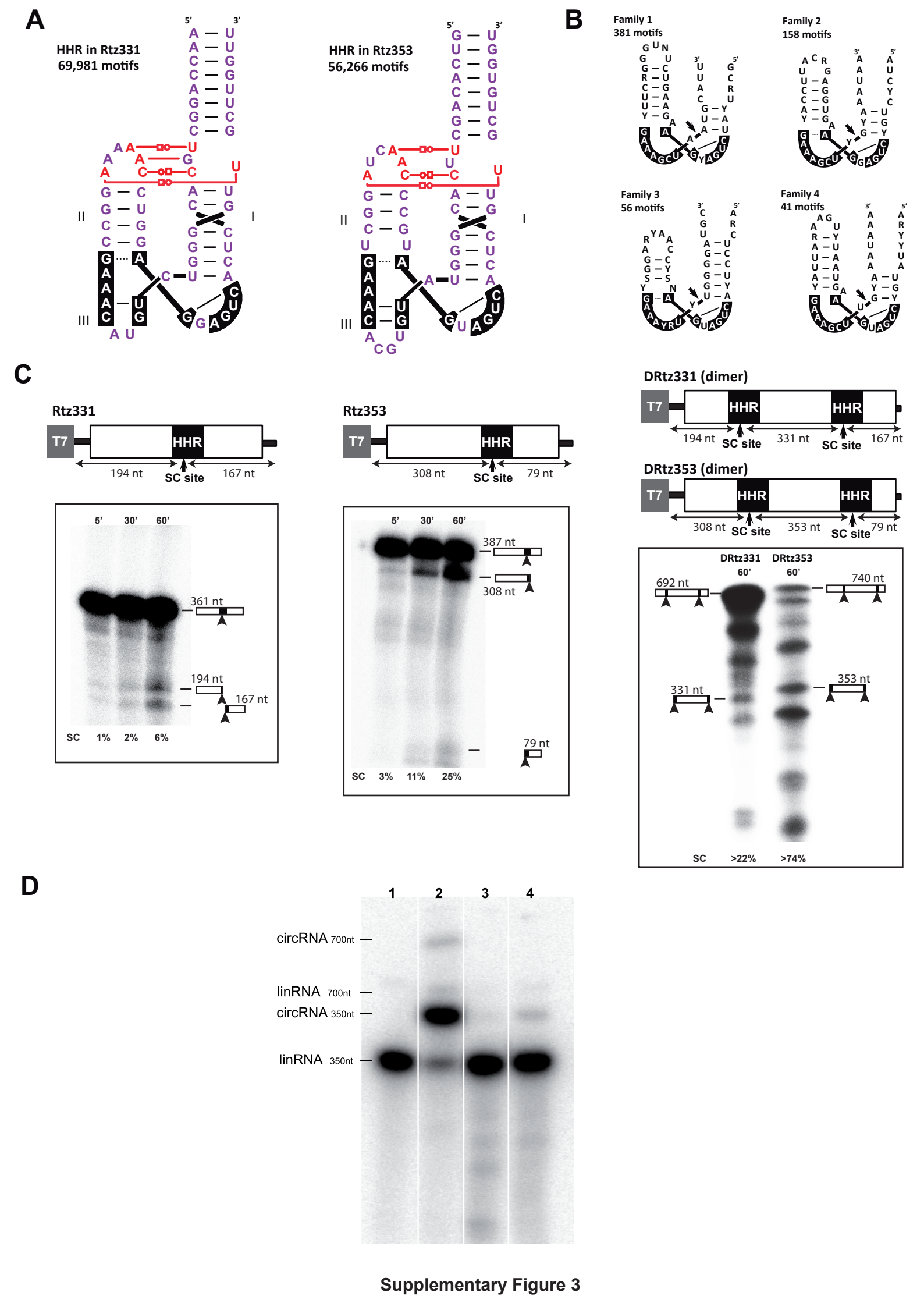
B



Supplementary Figure 2

Supplementary Figure 2

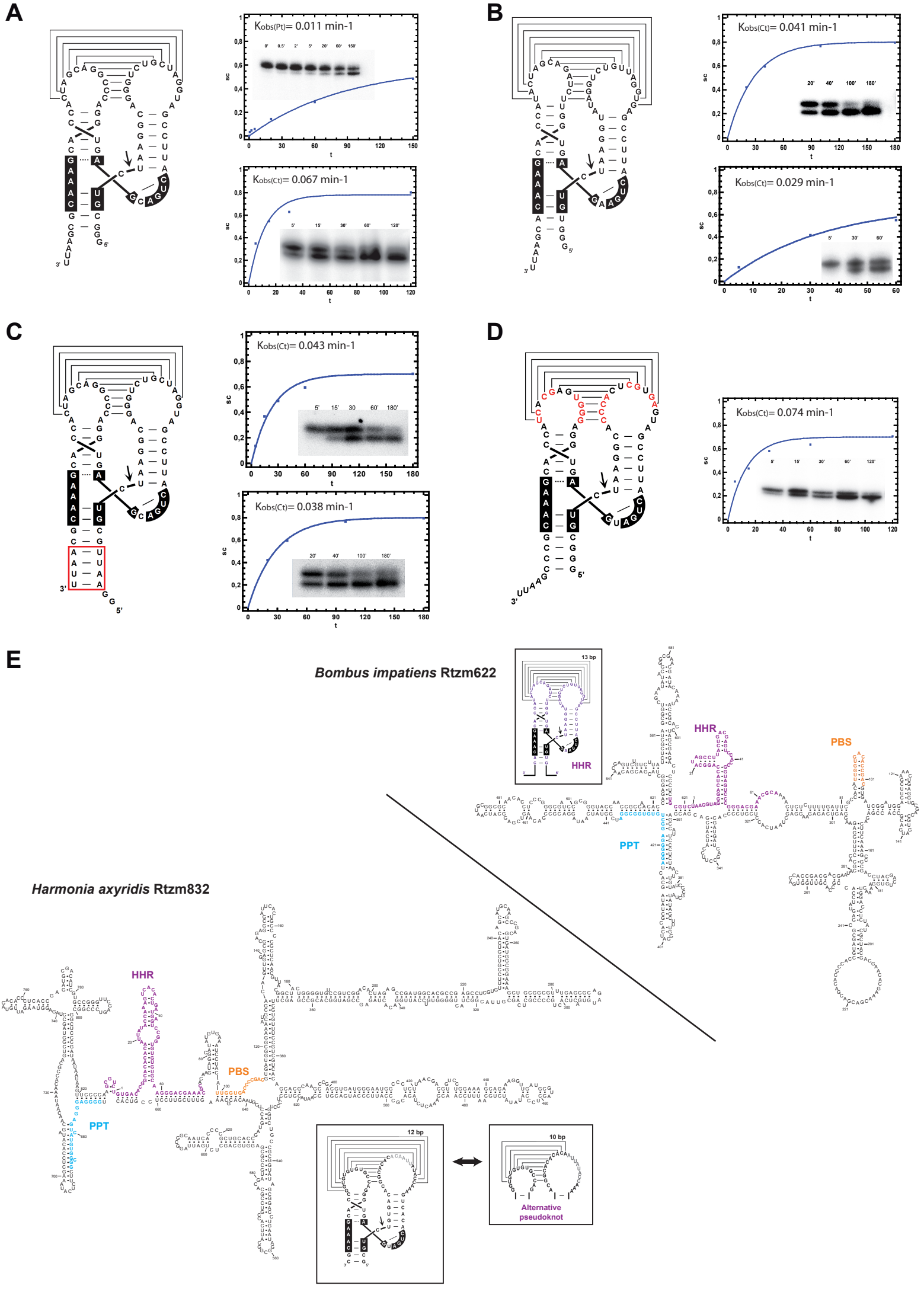
(A) A genomic copy of the mussel *M. galloprovincialis* retrozyme was cloned in a pBlueScript vector. Schematic representation of the clone and the expected sizes of the fragments after HHR self-cleavage are indicated and autoradiography of a run-off transcription at different times in the presence of [α-32P]-ATP. Cleavage of the whole RNA of the retrozyme reaches almost completion (98%). (B) Sequence alignment of selected cloned circRNAs from the male gonads of a *M. galloprovincialis* mussel. The first positions of the sequences correspond to the self-cleavage site of the retrozymes. Highest sequence homology occurs among either the ribozyme motifs or the central region of the retrozymes (positions ~100-150 nt).



Supplementary Figure 3

Supplementary Figure 3

(S) The two families of canonical type I HHRs found in the genome of the Mexican axolotl. The ribozyme motif depicted at the left occurs in the retrozymes of ~330 bp whereas the one at the right is found in the retrozymes of ~350 bp. (B) Consensus sequences of four representative examples of minimal HHRs of PLE and Terminon retrotransposons detected in the genome of the axolotl. (c) Sequence alignment of a couple of retrozymes from the axolotl genome corresponding to each family, Rtzm 330 and 350). Sequence homology is mostly restricted to the HHR and surrounding regions (the first and last 100 bp of the retrozymes). (C) Autoradiography of run-off transcriptions of genomic retrozymes of the mexican axolotl at different times in the presence of [α-32P]-ATP. Left (Rtz331) and center (Rtz353) panels correspond to monomeric motifs with the HHR located in the indicated positions, whereas right panel corresponds to dimeric constructs of these two retrozymes (DRtz331 and DRtz353). Self-cleavage of the sequences during 1h transcription reaches a maximum of 6% and 25% for the monomeric HHRs, whereas dimeric constructs reached 22% and 74% self-cleavage for Rtz331 and Rtz353, respectively. The gel purified monomeric RNA resulting from double self-cleavage of Rtz350 was run either directly in a denaturing gel (lane 1), or after incubation with RtcB ligase in its corresponding buffer (lane2), or after incubation in RtcB buffer without the protein ligase (line 3) or after incubation in 40 mM Mg2+ buffer (lane 4). Circular RNAs can be readily detected in lane 2 (80% circularization) and lane 4 (3% self circularization). Minor fractions of linear and circular dimeric RNAs are also indicated (lane 2).



Supplementary Figure 4

Supplementary Figure 4

(A) A genomic pknot-HHR from the *Dufourea novaeangliae* bee (left) and two representative kinetic analysis of self-cleavage under post-transcriptional (top, right) and co-transcriptional (bottom, right) conditions. (B) A genomic pknot-HHR from the *Bombus impatiens* bumblebee (left) and two kinetic analyses under co-transcriptional conditions. (C) A mutant version of the pknot-HHR of *D. novaeangliae* bee with an elongated helix III (left, the 4 extra base pairs are indicated with a red box) and two kinetic analyses under co-transcriptional conditions. (D) A mutant version of the pknot-HHR of *D. novaeangliae* bee with 16 point mutations at the pseudknot interaction (in red). The observed constant of cleavage is indicated on the top each graph. For more details, see Materials and methods section. (E) Two representative examples of the secondary structure of minimum free energy of circRNAs derived from genomic retrozymes from the bumblebee *Bombus impatiens* (upper panel) and the Asian ladybug *Harmonia axyridis* (lower pannel). Sequences corresponding to the hammerhead ribozyme (HHR), Primer binding site (PBS) and Polypurine tract (PPT) motifs are shown in purple, orange and blue, respectively. The structure of the HHR motifs are shown in the insets. The two possible pseudonknots interactions detected for the *H. axyridis* HHR are also indicated (lower insets).