Processing of insect retrotransposons by self-cleaving ribozymes

Dana J. Ruminski[1,4], Chiu-Ho T. Webb[1,4], Nathan J. Riccitelli[2] and Andrej Lupták[1-3]

Departments of [1]Molecular Biology and Biochemistry, [2]Chemistry and [3]Pharmaceutical

Sciences, University of California, Irvine, California, USA.

[4]These authors contributed equally to this work.

Correspondence should be addressed to A.L. (aluptak@uci.edu)

Abstract

We show that several classes of insect non-LTR retrotransposons harbor self-cleaving ribozymes of the HDV family at their 5′ termini. In *Drosophila* the R2 ribozymes exhibit highly differential *in vivo* expression and robust *in vitro* activity, modulated by an upstream sequence originating from the insertion site. Our data suggest a role for self-cleaving ribozymes in co-transcriptional processing of retrotransposons with implications for downstream events, including translation and retrotransposition.

Retrotransposons are mobile genetic elements that utilize reverse transcriptase activity to propagate *via* a RNA intermediate in many eukaryotic genomes. The large fraction of eukaryotic genomes composed of retrotransposons coupled with contributions to ectopic recombination and novel gene regulatory elements suggest potential for major evolutionary influence[1]. Autonomous retrotransposons are divided into long terminal repeat (LTR) and non-LTR or LINE (long interspersed repeat elements) elements. SINEs (short interspersed elements) are non-autonomous elements thought to be mobilized primarily by non-LTR type elements[2]. LTR elements have promoters within the LTRs[2], however, the promoters for non-LTR elements are less well conserved and they can even be transcribed as introns of larger host transcripts[3]. The non-reverse transcriptase (RT) structural and catalytic proteins encoded by the two autonomous types of elements are also quite distinct[4].

One of the best-characterized non-LTR retrotransposons is R2, which inserts site-specifically into the 28S rDNA of most insects[5]. R2 was first identified as an insertion sequence in *Drosophila melanogaster*[6] and transcripts of interrupted rRNA genes were shown to be processed co-transcriptionally at the boundary of the 28S transcript and 5′ and 3′ ends of the intervening sequence[7]. R2 is co-transcribed with the 28S rRNA precursor by RNA polymerase I[3]. Across thirteen *Drosophila* species, the 5′ ends of R2 elements contain conserved features predicted to form a folded RNA secondary structure[3]. The mechanism of R2 retrotransposition has been studied in depth, leading to a general model for non-LTR reverse transcription and integration in which the endonuclease-catalyzed cleavage of the target DNA backbone is followed by target-primed reverse transcription (TPRT), second strand synthesis by the RT, and insertion of the newly synthesized cDNA into the genome[8]. Although much of this model has

been validated *in vitro*, the mechanism of co-transcriptional processing of these elements has been unknown.

Recently we have used structure-based searches to identify the hepatitis delta virus (HDV) family of self-cleaving ribozymes in many organisms[9]. These ribozymes fold into an intricate double-pseudoknot structure composed of five based-paired regions joined by single-stranded regions[10]. We found that some of the sequences map to predicted RT genes in the purple sea urchin *Strongylocentrotus purpuratus* and one ribozyme maps to the 5′ terminus of the RTE retrotransposon in the African mosquito *Anopheles gambiae*[9]. These results suggested that self-cleaving ribozymes might function in retrotransposition. However, the predicted secondary structure of these sequences differs somewhat from structures of canonical HDV-like ribozymes: some examples of the sea urchin ribozymes have shorter P1 regions than was previously observed and the mosquito RTE ribozyme (drz-Agam-2) contains an unusually large peripheral domain in the J1/2 region[9].

To test more broadly for the association of HDV-like ribozymes with retrotransposons, we performed motif searches[11] that allowed for variable P1 helices and large inserts in the J1/2 region of the ribozyme. We searched ESTs, RT-containing mRNAs and sequences defined as repeats by RepeatMasker in various insect genomes. The expanded search criteria allowed for identification of many ribozyme candidates, a subset of which mapped to the 5′ termini of known retrotransposons (**Table 1, Supplementary Table 1, and Supplementary Fig. 1**). These included additional *A. gambiae* RTE elements, and R2, Bilbo and Baggins non-LTR retrotransposons in various *Drosophila* species. We found that the conserved regions at the 5′ end of R2[3] form the catalytic core of the putative ribozyme double-pseudoknot structure (**Fig.**

1a), while the less-conserved sequences map primarily to the peripheral domains (J1/2 and P4-L4).

To investigate the activity of the *Drosophila* ribozymes, we amplified the genomic regions containing the 5′ termini of the R2 and Baggins elements from several *Drosophila* species and tested their Mg^{2+}-dependent self-cleavage *in vitro*. Similar to the *A. gambiae* ribozymes[9], the *Drosophila* RNAs showed robust self-cleavage kinetics at ambient temperature and physiological Mg^{2+} concentrations (**Fig. 1b and 1c**), demonstrating that these structures are *bona fide* ribozymes. In the R2 ribozymes, the cleavage rate constants varied among isolates, which differed mostly in composition of the leader sequence just upstream of the cleavage site and the J1/2 peripheral region. This suggests that folding of the sequence surrounding the cleavage site and of the peripheral domain affect the formation of active ribozymes. In HDV ribozymes the sequence upstream of the cleavage site has previously been shown to influence proper folding of the ribozymes into active conformations[12]. When this leader sequence has the capacity to extend the P1 helix of the ribozyme, it prevents formation of the P1.1 region of the the active site, resulting in decreased activity. Our data correlate well with this hypothesis: we observe significantly faster self-cleavage kinetics in identical ribozymes preceded by leader sequences that cannot extend the P1 helix and slower kinetics in constructs that can form the longer P1 (**Fig. 1d**). Interestingly, a typical insertion site for the R2 element in the 28S rRNA gene[6] creates a leader sequence that supports fast ribozyme self-cleavage. We speculate that efficient co-transcriptional 5′ processing of the element is beneficial to the retrotransposon and that insertion into the typical site in the 28S rDNA provides an evolutionary advantage to R2 retrotransposon propagation.

The Bilbo and Baggins retrotransposons and their 5′ ribozymes were found in only a subset of the *Drosophila* species. The ribozyme-terminated Baggins retrotransposons are found in *D. melanogaster*, *D. sechellia*, *D. simulans*, *D. ananassae*, *D. grimshawi*, and *D. persimilis*. These ribozymes have predicted structures similar to the R2 RNAs, with large peripheral domains in the J1/2 region, but longer P4-L4 regions (**Table 1 and Supplementary Fig. 1**). Their *in vitro* activity is also variable and the sequences we tested had somewhat lower cleavage rate constants than the R2 ribozymes. Nonetheless, their activity was robust and, as in the case of other HDV-like ribozymes[9,13], showed temperature and Mg^{2+} dependence.

To investigate the *in vivo* activity of the ribozymes, we analyzed total RNA isolated from several life stages of *D. melanogaster* (W1118). Similar to *A. gambiae*, where the ribozymes are differentially expressed[9], we found that *D. melanogaster* R2 ribozymes were highly expressed in adult females and pupae and less so in larvae and adult males (**Fig. 2a**). The expression difference between adult males and females is striking because the level of ribosomal RNA expression is the same. This suggests that different rDNA loci are transcribed in males and females or that R2 elements inserted in non-rDNA loci are expressed in females. Expression analysis[14,15] and electron microscopy[7] have previously shown that the R2 retrotransposons are processed co-transcriptionally at the 5′ termini. Together these data show that the R2 retrotransposons and their terminal ribozymes are highly expressed and active *in vivo*, leading to a model of R2 RNA production involving co-transcriptional cleavage by an HDV-like ribozyme (**Fig. 2b**).

The role of self-cleaving ribozymes beyond 5′ terminal processing is speculative, however two potential functions have previously been proposed for this region. The 5′ terminal structure of the R2 element harbors a short coding region just upstream of the RT coding region[3].

The start and stop codons map to the P4 helix of the self-cleaving ribozyme (**Fig. 1a**), suggesting that the ribozyme structure may affect translation of the R2 transcript. The second function may be in insertion of the R2 cDNA. The R2 RT enzyme efficiently switches templates upon reaching the 5′ end of the R2 RNA, presumably to facilitate insertion of the element into a new genomic site[16]. This *trans*-templating reaction by the R2 protein is significantly more efficient when the 5′ end of the RNA template consists of a 5′ hydroxyl group, as compared to a methylguanosine cap analog or a triphosphate[17]. The 5′ product of the self-scission reaction in HDV-like ribozymes is a hydroxyl group, thus ribozyme-terminated RNA may be a superior substrate for *trans*-templating by the R2 protein and may be beneficial for overall retrotransposon insertion efficiency. The model we describe here suggests that these elements have linked transcription, translation, and transposition in such a way that templates with ribozyme-terminated 5′ ends are efficient retrotransposons. The mechanism through which the 5′ end is retained in these non-LTRs is distinct from LTR elements and provides protection from 5′ end erosion in actively propagating elements.

Our results show that multiple insect non-LTR retrotransposons harbor self-cleaving ribozymes that are used to liberate their 5′ termini from upstream transcripts. This work explains observations made over 25 years ago that showed that the *Drosophila* R2 retrotransposons are processed co-transcriptionally and points to several possible roles of self-cleaving ribozymes in retrotransposition. Our finding supports a general trend indicating that self-cleaving ribozymes are associated with many transposable elements, including the human LINE1 non-LTR retrotransposon[13], and satellites in newts[18], fungi[19] and schistosomes[20].

Acknowledgments

AUTHOR CONTRIBUTIONS

A.L. and D.J.R. designed experimental strategies; D.J.R., C.-H.T.W. and N.J.R. performed experimental analysis and interpreted data; A.L. and D.J.R. interpreted data and wrote the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Figure legends

Figure 1 *D. melanogaster* drz-Dmel-1-1 ribozyme. (**a**) *Drosophila* R2 ribozyme consensus secondary structure. Core elements are colored by region corresponding to the HDV ribozyme[10]. Boxed nucleotides are start and stop codons found upstream of the R2 coding region[3]. Green arrow points to the cleavage site. Consensus is based on *in vitro* active R2 ribozymes (**Table 1**). (**b**) *In vitro* self-cleavage activity of ^{32}P-labeled ribozyme incubated in 10 mM MgCl$_2$ at 37°C. (**c**) Semi-logarithmic graph of *in vitro* self-cleavage reactions at indicated conditions. (**d**) Potential extension of the P1 helix with the leader sequence. Alternative pairing between the leader sequence (black) and the P1.1 region (gray) is shown in red. Rate constants listed were measured in 10 mM MgCl$_2$ at 37°C.

Figure 2 R2 retrotransposon activity in *Drosophila*. (**a**) Expression of drz-Dmel-1 ribozymes relative to GAPDH mRNA (logarithmic scale) measured using RT-qPCR of total RNA isolated from indicated developmental stages. RT-qPCR primers were designed to amplify all ribozyme-containing R2 elements. All data are average values ± average deviations. (**b**) Proposed model of R2 RNA processing and expression.

Methods

All ribozymes constructs were amplified from their respective genomes using primers listed below. Naming of the ribozymes follows the rules proposed previously[9]. Briefly, the name of the ribozyme was derived from the binomial nomenclature for the species (drz = delta-like ribozyme, *Drosophila ananassae* – Dana). Among the fruit fly ribozymes, family 1 (e.g. drz-Dmel-1) corresponds to R2 (or R2B) ribozymes and family 2 (e.g. drz-Dmel-2) corresponds to Baggins ribozymes.

Drosophila melanogaster (strain: W1118)

drz-Dmel-1-1

Ribozyme construct:
5′aagcgctggtcaacggcgggagtaactatgactgtcttaaGGGGAGTCATGGGGTATTTGAGAGCAGAGG
GGGAGTATTCTTCTGTAATTCGTAAGTCATATCATATGATGTGCGGAAGGGGAATTT
TACTCTGTAACTCACAAGTCTCTCCTTTACTCAAGTCGACTCAAAACCTCCTCGTGGT
GGTCCCCGGTAATGCTAAACTTGTTTAGCAGCTAATTTGAGCGGCaaaactt

AL506: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGAAGCGCTGGTCAACGGCGGG
AL507: 5′ AAGTTTTGCCGCTCAAATTAGCTGCTA
AL508 (inhibitor): 5′ GACTCCCCTTAAGA

drz-Dmel-1-2

Ribozyme construct:
5′ggcgagctgatcactgattggggtgactgcgcacgagttGGGGATCATGGGGTATTTGAGAGCAGAGGG
GGAGTATTCTTCTGTAATTCGTAAGTCATATCATATGGTGTGCGGAAGGGGAATTTT
ACTCTGTAACTCACAAGTCTCTCCTTTACTCAAGTCGACTCAAAACCTCCTCGTGGTG
GTCCCCGGTAATGCTAAACTCGTTTAGCAGCTAATTTGAGCGGCaaaac

AL527: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGGGCGAGCTGATCACTGATTGG
AL528: 5′ GTTTTGCCGCTCAAATTAGCTGCTA
AL529 (inhibitor): 5′ GATCCCCAACTCG

drz-Dmel-2-1

Ribozyme construct:
5′gttttgtctccggattctccaattgttatttatttataatGGCCGCCATGACAGTAGGTATCACAAGGGGATCA
ACGCGCCACCACTACTGGAAGTGCAACTACACTCTCCACGCGAGGGCGGCTGGGAA
CAGGCTCTCAGTTAGGTCATGTCTCTGCTAAGAGTGCTGGCTAATCATAGTTGggct

AL512: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGGTTTTGTCTCCGGATTCTCCAA
AL565: 5′ AGCCCAACTATGATTAGCCAGC
AL514 (inhibitor): 5′ GGCGGCCATTAT

drz-Dmel-2-2

10

Ribozyme construct:

5′cacgatatctacggatctcgcggtaacttaaaccaaaaaaGGCCGCCACGACAGTAGATATCACAAGGGG
ATCAACGCGCCACCAACGGTGGTACGCGCCGTAATTGTGAAACACTACTGGACGTG
CAACTACACTCTCCACGCGAGGGCGGCTGGAAACAGGCTTTTAGTAAGGTCATGTA
ACTGCTAAGAGTGCTGGCTAATCGTAGTTGggctgg

AL515: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGCACGATATCTACGGATCTCG
AL516: 5′ CCAGCCCAACTACGATTAGCCA
AL517 (inhibitor): 5′ GGCGGCCTTTTT

drz-Dmel-2-3

Ribozyme construct:

5′atatatttatattatttgccatgcccttaagccccgaaaaGGCCGCCACGCCTATAGGTATCACAAGGGGAT
CAACGCGCCACCAACGGTGGTACACACCGTAACTGTGACATACTACTTGAAGTGCA
ACTACACTCTCCACGCGAGGGCGGCTGGAAACAGGCTCTTAGTTAGGACCCTGCTA
AGAGTGCTGGCTAATCGTAGTTGggctgg

AL518: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGATATATTTATATTATTTGCCATGCCCT
AL519: 5′ CCAGCCCAACTACGATTAGCCAG
AL520 (inhibitor): 5′ GGCGGCCTTTTC

Drosophila simulans (strain: (Matzkin) Isosancruz-19)

drz-Dsim-1

Ribozyme construct:

5′agtaagcgcgggtcaacggcgggagtaactatgactctctGAGGGATCTGGGGTAATTGCGAGCAGAGGG
GGAGTATTTTTCTGTAATTCGTAAGTCATATCATATGGTGTGCGGAAGGGGAATTTT
ACTCTGTAACTCACAAGTCTCTCCTTTACTCAAGTCGACTCAAAACCTCCTCGTGGTG
GTCCCCGGTAATGCTAAACTTGTTTAGCAGCTAATTTGAGCGGCaaaaactt

AL566: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGAGTAAGCGCGGGTCAACGGCG
AL567: 5′ AAGTTTTTGCCGCTCAAATTAGCTGCT
AL568 (inhibitor): 5′ GATCCCTCAGAGAG

Drosophila ananassae (strain: (BGS) SB18.8C)

drz-Dana-1

Ribozyme construct:

5′taagcgcgggtcaacggcgggagtaactatgactctctttGGAGAATATGGATTTGATTGTGCAGAGGGGG
TGCTATACCGTAACTCGTAAGCCATGCAATCAGATCAAGTCGACTCAAAACCTCCTC
GTGGTATTCTCTGGGTGCCAGTATTTACTGGTAGCTGATTTGAGCGGCgaaag

AL581: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGTAAGCGCGGGTCAACGGC

AL581: 5′ CTTTCGCCGCTCAAATCAGCT
AL583 (inhibitor): 5′ ATATTCTCCAAAGAGA

Drosophila sechellia (strain: (BSC) 3529 line 15)

<u>drz-Dsec-1</u>

Ribozyme construct:
5′agtaagcgcgggtcaacggcgggagtaactatgactcttaGGGGATCAGGGGTAATTGCGAGCAGAGGGG
GAGTATTTTTCTGTAATTCGTAAGTCATATCATATGGTGTGCGGAAGGGGAATTTTA
CTCTGTAACTCACAAGTCTCTCCTTTACTCAAGTCGACTCAAAACCTCCTCGTGGTGG
TCCCCGGTAATGCTAAACTTGTTTAGCAGCTAATTTGAGCGGCaa

AL798: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGAGTAAGCGCGGGTCAACGGC
AL799: 5′ TTGCCGCTCAAATTAGCTGCTA
AL800 (inhibitor): 5′ GATCCCCTAAGAG

Drosophila persimilis (strain: CC23)

<u>drz-Dper-1</u>

Ribozyme construct:
5′aagcgcgggtcaacggcgggagtaactatgactctcttaaGGAAGATATGGATCTGAACAATAGCGTAGA
AGGGGAGTCATTCCGTAATTCGTAAATCGTAAAAATTAGATCAAGTTGATTCAAGAC
CTCCTCGTGGTATCTTCTGGATGCTATTAGACTAAAGTTCTTTTGGTCTAATAGTAAC
TAACTTGAACagcgaaa

AL645: 5′ TTCCCGCGAAATTAATACGACTCACTATAGGGAGAAGCGCGGGTCAACGGCG
AL646: 5′ TTTCGCTGTTCAAGTTAGTTACTATT
AL644 (inhibitor): 5′ ATCTTCCTTAAGAGA

<u>drz-Dper-2</u>

Ribozyme construct:
5′tcagcttgtccgatattgtagtagtttttaagttgacccaGGCCGTCAGCCATCGGGTGCTACAGGGGGACA
AACGGATGACGGGTCGGCTCGCACAACTGTGAACTCGGAACCCGCCGTAACTGTAC
ACACCGGTGGAAGTGCGACTATACTCTCCCCGTGAGGGCGGCTGGAAACAGGTCCT
GGCACGGTCATGTCTCTGCCAGGATTCTGGCGAATGGTAGTCGggcg

AL633: 5′TTCCCGCGAAATTAATACGACTCACTATAGGGAGTCAGCTTGTCCGATATTGTAGTAG
AL634: 5′ CGCCCGACTACCATTCGCC
AL635 (inhibitor): 5′ GACGGCCTGGG

Transcription, kinetic analysis of ribozyme self-cleavage, RNA extraction, and RT-qPCR were performed as described previously[9]. Primers used for qPCR are listed below.

drz-Dmel-1: forward - 5′ CGACTCAAAACCTCCTCG TGGT, reverse - 5′ TAGCCGCTGCGTTTGGTTCATA
GAPDH: forward - 5′ GGACTCACGGTCGTTTCAAG, reverse - 5′ GCCTTGTCAATGGTGGTGAA

References.

1. Bourque, G. Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr Opin Genet Dev* **19**, 607-12 (2009).
2. Mourier, T. & Willerslev, E. Retrotransposons and non-protein coding RNAs. *Brief Funct Genomic Proteomic* **8**, 493-501 (2009).
3. George, J.A. & Eickbush, T.H. Conserved features at the 5 end of Drosophila R2 retrotransposable elements: implications for transcription and translation. *Insect Mol Biol* **8**, 3-10 (1999).
4. Eickbush, T.H. & Jamburuthugoda, V.K. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res* **134**, 221-34 (2008).
5. Jakubczak, J.L., Burke, W.D. & Eickbush, T.H. Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proc Natl Acad Sci U S A* **88**, 3295-9 (1991).
6. Roiha, H., Miller, J.R., Woods, L.C. & Glover, D.M. Arrangements and rearrangements of sequences flanking the two types of rDNA insertion in D. melanogaster. *Nature* **290**, 749-53 (1981).
7. Jamrich, M. & Miller, O.L., Jr. The rare transcripts of interrupted rRNA genes in Drosophila melanogaster are processed or degraded during synthesis. *EMBO J* **3**, 1541-5 (1984).
8. Kurzynska-Kokorniak, A., Jamburuthugoda, V.K., Bibillo, A. & Eickbush, T.H. DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon. *J Mol Biol* **374**, 322-33 (2007).
9. Webb, C.-H.T., Riccitelli, N.J., Ruminski, D.J. & Luptak, A. Widespread Occurrence of Self-Cleaving Ribozymes. *Science* **326**, 953 (2009).
10. Ferre-D'Amare, A.R., Zhou, K. & Doudna, J.A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567-74 (1998).
11. Riccitelli, N.J. & Luptak, A. Computational Discovery of Folded RNA Domains in Genomes and in vitro Selected Libraries. *In press* (2010).
12. Chadalavada, D.M., Knudsen, S.M., Nakano, S. & Bevilacqua, P.C. A role for upstream RNA structure in facilitating the catalytic fold of the genomic hepatitis delta virus ribozyme. *J Mol Biol* **301**, 349-67. (2000).
13. Salehi-Ashtiani, K., Luptak, A., Litovchick, A. & Szostak, J.W. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* **313**, 1788-92 (2006).
14. Kidd, S.J. & Glover, D.M. Drosophila melanogaster ribosomal DNA containing type II insertions is variably transcribed in different strains and tissues. *J Mol Biol* **151**, 645-62 (1981).
15. Long, E.O., Collins, M., Kiefer, B.I. & Dawid, I.B. Expression of the ribosomal DNA insertions in bobbed mutants of Drosophila melanogaster. *Mol Gen Genet* **182**, 377-84 (1981).
16. Bibillo, A. & Eickbush, T.H. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. *J Mol Biol* **316**, 459-73 (2002).
17. Bibillo, A. & Eickbush, T.H. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J Biol Chem* **279**, 14945-53 (2004).

18. Epstein, L.M. & Gall, J.G. Self-cleaving transcripts of satellite DNA from the newt. *Cell* **48**, 535-43 (1987).
19. Saville, B.J. & Collins, R.A. A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell* **61**, 685-96 (1990).
20. Ferbeyre, G., Smith, J.M. & Cedergren, R. Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol Cell Biol* **18**, 3880-8 (1998).

Figure 1.

a

b

Time

sec | min | hr

0 6 12 20 30 40 50 1 2 4 8 16 30 1 2 3

Precursor

3′ Product (ribozyme)

c

◆ 10mM MgCl$_2$, 37 ºC   ● 1mM MgCl$_2$, 37 ºC

◇ 10mM MgCl$_2$, 22 ºC   ○ 1mM MgCl$_2$, 22 ºC

Fraction Intact

Time (min)

d

Ribozyme

drz-Dmel-1-1   drz-Dsec-1   drz-Dsim-1   drz-Dmel-1-2

Rate constants (hr^{-1})

72   52   4.2   0.30

Figure 2.

a

b
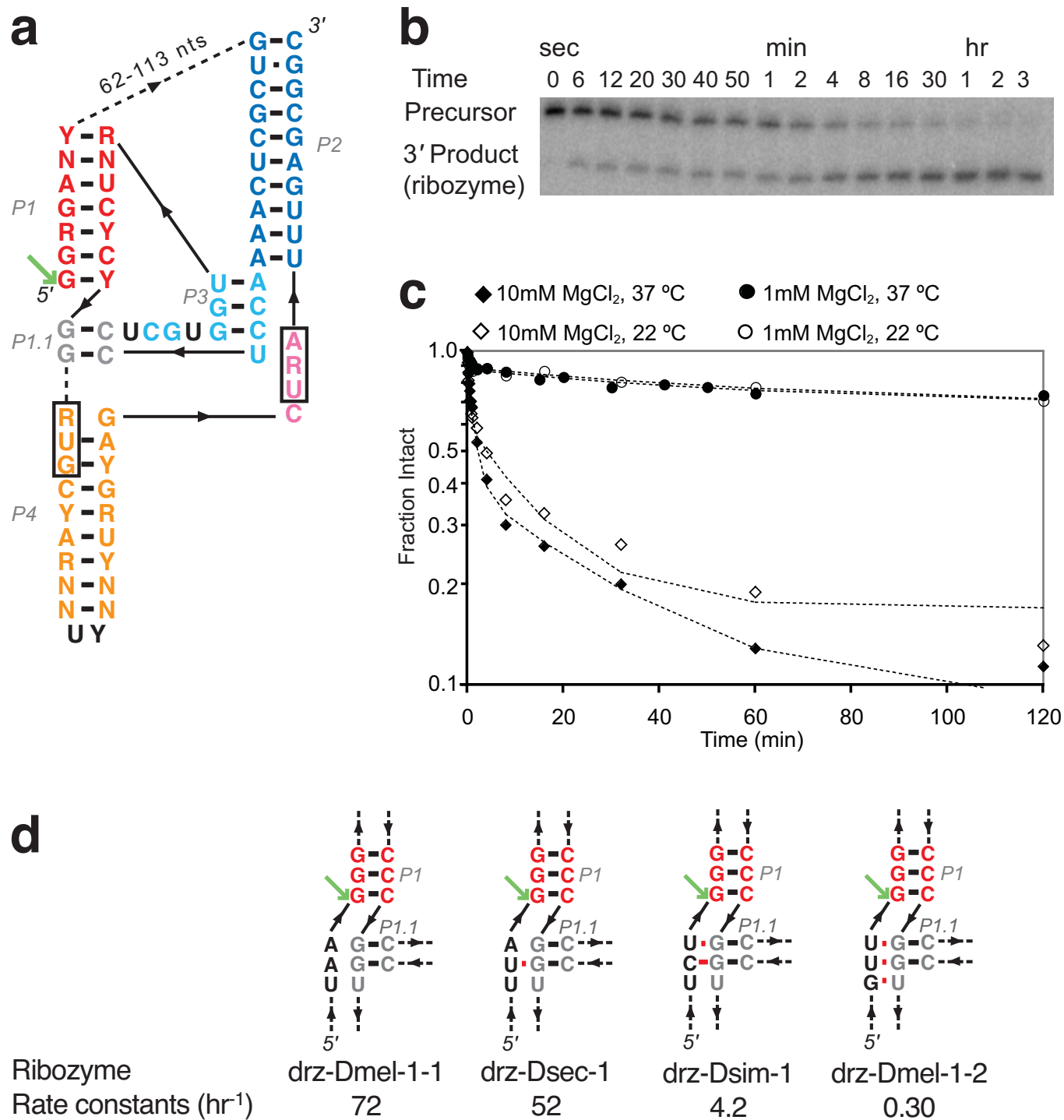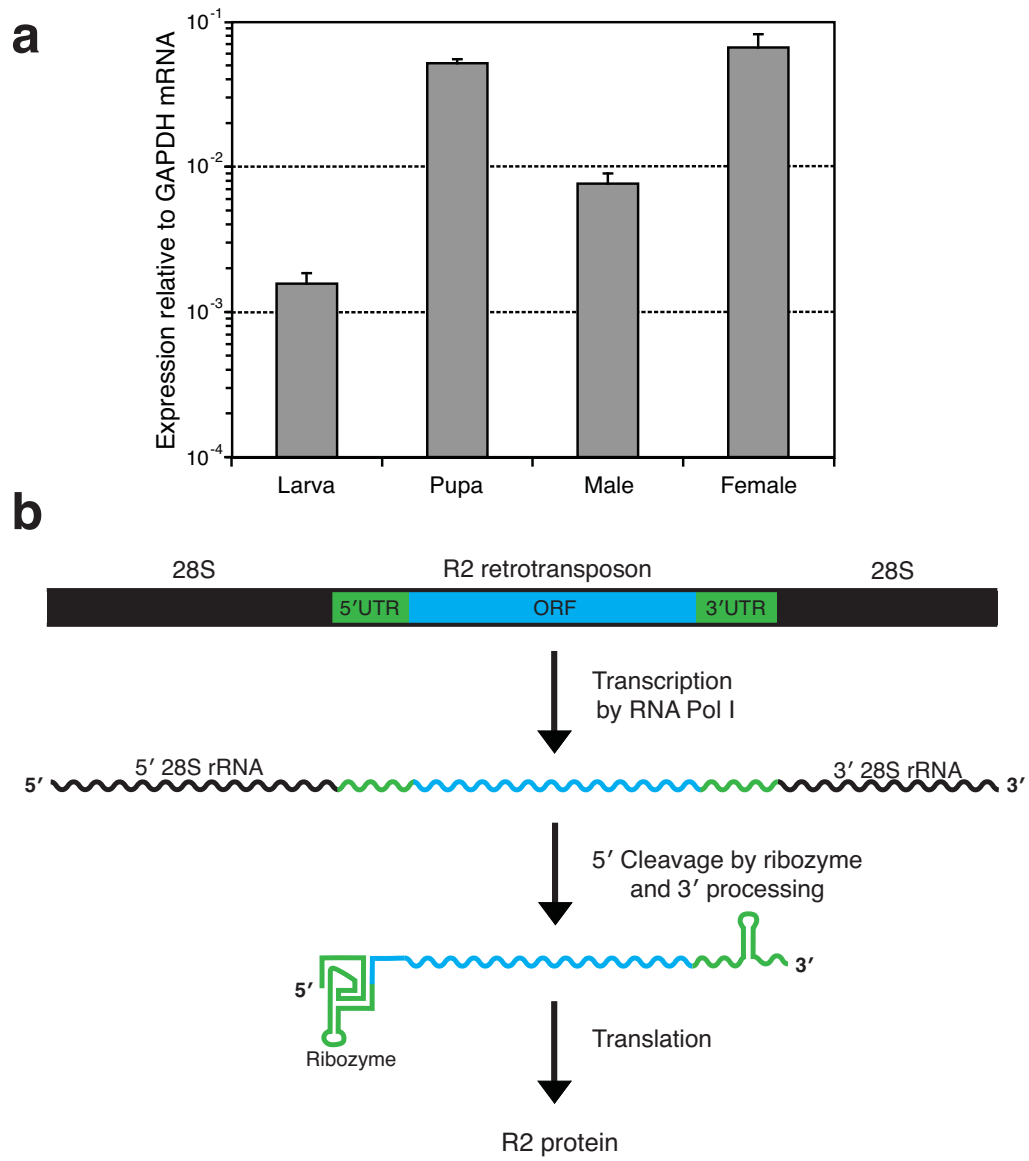
Table 1 Retrotransposon ribozymes with confirmed *in vitro* activity

	Genomic locus (strand)	P1	J1/2 Length	P2	P3	L3	P3'	P1'	J1.1/4	P4 Length	J4/2	P2'	Ribozyme Name	Rate Constant/hr[-1]	Number of copies in reference genome
R2															
D. melanogaster	chrUextra: 5231803 (-)	GGGGAGUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	drz-Dmel-1-1	71 ± 31	39
	chrXHet: 198910 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	drz-Dmel-1-2	0.30 ± 0.01	
D. sechelllia	super_700: 5309 (-)	GGGGAUC	112	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	drz-Dsec-1	~ 50	12
D. simulans	chrX_random: 649216 (-)	GGGGAUC	112	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	drz-Dsim-1	~ 4	11
D. ananassae	scaffold_13163: 86767 (+)	GGAGAAU	62	GUCGACUCAAA	ACC	UCCUCGU	GGU	AUUCUCU	GGG	18	GCUGA	UUUGAGCGGC	drz-Dana-1	> 60	3
R2B															
D. persimilis	super_2450: 4278 (-)	GGAAGAU	69	UGAUUCAAG		ACC	UCCUCGU	GGU	AUCUUCU	GGA	35	ACUAA	CUUGAACA	drz-Dper-1	22 ± 3	1
Baggins															
D. melanogaster	chrU: 1256859 (-)	GGCCGCC	50	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGGAA	39	GCUAAU	CAUAGUUG	drz-Dmel-2-1	~ 1	12
	chrU: 2434288 (+)	GGCCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	drz-Dmel-2-2	11 ± 2	
	chr3LHet: 191774 (+)	GGCCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	34	GCUAAU	CGUAGUUG	drz-Dmel-2-3	0.19 ± 0.03	
D. persimilis	super_43: 160102 (+)	GGCCGUC	96	CGACUAUA	CUC	UCCCCGU	GAG	GGCGGCU	GGAAA	37	GCGAA	UGGUAGUCG	drz-Dper-2	1.1 ± 0.1	1
RTE															
A. gambiae	chrX: 7606912 (+)	GUUCUGU	46	ACUCAAU	ACG	UCCUCGU	CGU	ACAGAAC	GGUAA	81	GCUAA	AUUGAGA	drz-Agam-2-2	ref. 11	36

All kinetics were measured in 10 mM MgCl$_2$ at 37°C, except for drz-Dper-2, which was measured in 1 mM MgCl$_2$ at 22°C .

Supplementary Figure 1.
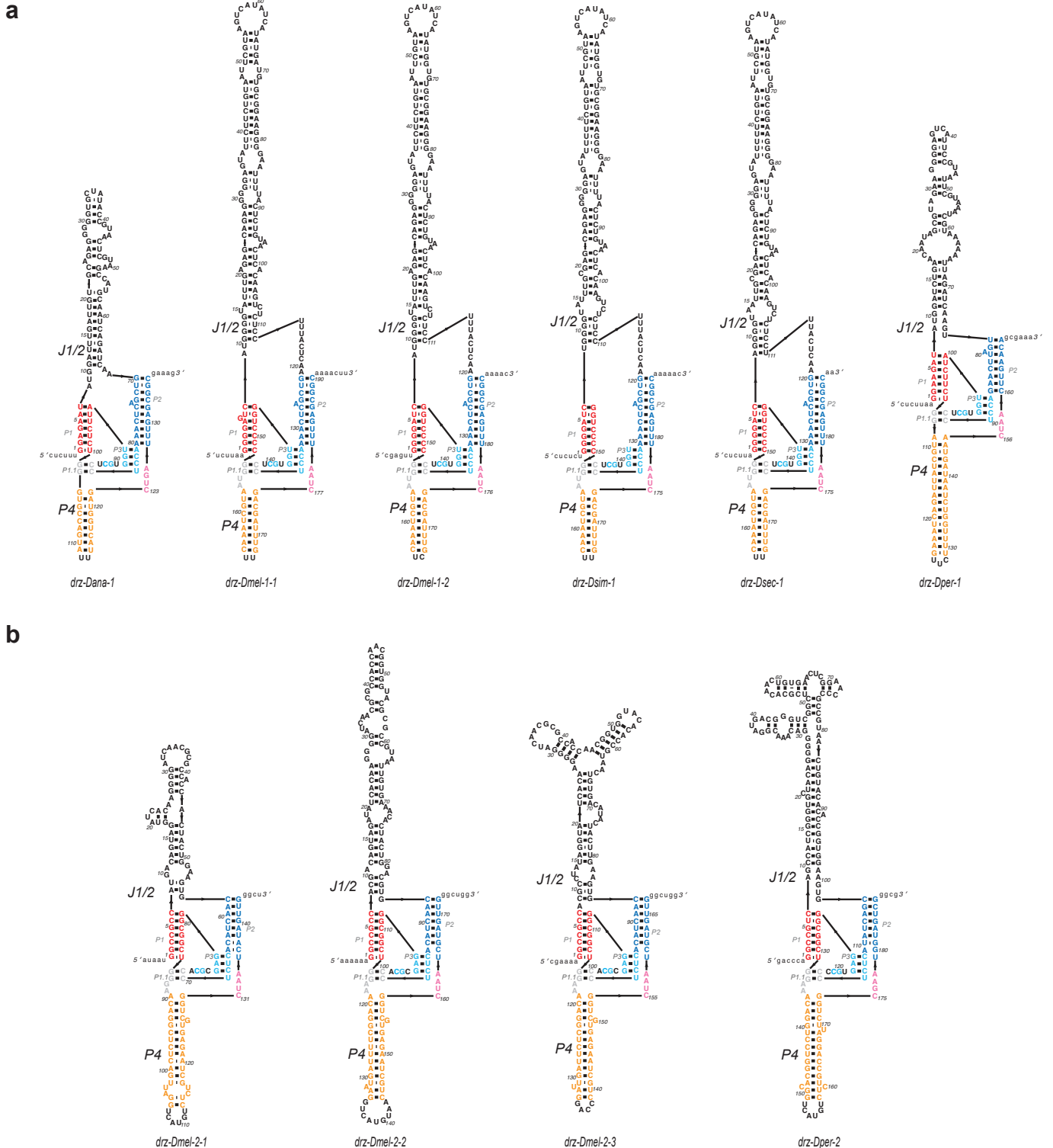
Supplementary Figure 1. Proposed secondary structures of HDV-like ribozymes identified in *Drosophila* retrotransposons. Paired elements of the structurally conserved cores are grouped by color. The peripheral J1/2 domains are shown in black. Predicted base-pairing is indicated by short horizontal lines. Lines with arrows indicate connections and do not correspond to any sequence. *Drosophila* R2 (drz-Dper-1 is from R2B) (**a**) and Baggins (**b**) ribozymes are shown with the ribozyme names indicated below their secondary structures.

Supplementary table 1 Putative retrotransposon ribozymes

Organism	Retrotransposon	Genomic locus (strand)	P1	J1/2 Length	P2	P3	L3	P3'	P1'	J1.1/4	P4 Length	J4/2	P2'	Ribozyme Name
A. gambiae	RTE	chr2L:1325088 (+)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr2L:13681013 (-)	GCUCUGC	59	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr2L:5793525 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr2R:10640126 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr2R:42188194 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr2R:47201145 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr2R:48807223 (+)	GCUCUGC	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUAA	48	CCUUA	GGCGGAU	
A. gambiae	RTE	chr2R:52222381 (+)	GCUUUGC	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3L:12176742 (+)	GCUCUGC	58	AUUCCGAC	ACG	UCCUCGU	CGU	GUAGAGU	GCU	18	UCCAA	GCCUGGAU	
A. gambiae	RTE	chr3L:35802504 (+)	GCUCUGC	59	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	60	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3L:492623 (+)	GCUCUGC	59	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3L:4992577 (+)	GCUCUGC	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3L:5416869 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3R:17347810 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3R:24429181 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3R:26113924 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3R:36915833 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3R:40187311 (+)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3R:49291435 (+)	GCUCUGU	59	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chr3R:52939674 (+)	GCUCUUC	12	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUAGC	53	GCGGAU	GUUGGCAU	
A. gambiae	RTE	chrU:10705999 (+)	GCUCUGU	59	AUUCCAAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCA	
A. gambiae	RTE	chrU:10766686 (+)	GCUCUGC	59	UUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUAA	190	ACCGAC	UACAGAA	
A. gambiae	RTE	chrU:12052610 (+)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:19907466 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:19977002 (-)	GCUCUGU	59	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:21458910 (+)	GCUCUGU	56	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:2679251 (-)	GCUUUGC	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:28798192 (+)	GCUCUGU	19	CCCCAGUG	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	42	CCCAAC	CCCUUGGG	
A. gambiae	RTE	chrU:30448356 (-)	GCUCUGU	19	CCCCAGUG	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	42	CCCAAC	CCCUUGGG	
A. gambiae	RTE	chrU:35889172 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:50466685 (+)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:56500771 (-)	GCUCUGU	59	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GUUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:8826460 (-)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
A. gambiae	RTE	chrU:9504200 (+)	GCUCUGU	59	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GUUA	55	GCGGAU	GGUGGCAU	
A. gambiae	**RTE**	**chrX:7606912 (+)**	**GUUCUGU**	**46**	**ACUCAAU**	**ACG**	**UCCUCGU**	**CGU**	**ACAGAAC**	**GGUAA**	**81**	**GCUAA**	**AUUGAGA**	**drz-Agam-2-2**
A. gambiae	RTE	chrX:9302720 (+)	GCUCUGU	58	AUUCCGAC	ACG	UCCUCGU	CGU	GCAGAGU	GGUA	55	GCGGAU	GGUGGCAU	
D. melanogaster	R2	chrU:6635081 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrU:9469608 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrU:9660851 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:10741648 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:10784596 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:11156813 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:12028170 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:13744436 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:14001871 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:14308093 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	

Species		Location												
D. melanogaster	R2	chrUextra:15113533 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:15146783 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:15239032 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:16371793 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:1641391 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:17219920 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:18878225 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:21911713 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:23561203 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:2423598 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:25327518 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:3329827 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:5092520 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:5153848 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:5156753 (+)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	**R2**	**chrUextra:5231803 (-)**	**GGGGAGUC**	**113**	**GUCGACUCAAA**	**ACC**	**UCCUCGU**	**GGU**	**GGUCCCC**	**GGUAA**	**18**	**GCUAA**	**UUUGAGCGGC**	drz-Dmel-1-1
D. melanogaster	R2	chrUextra:7629486 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:8442033 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:9005695 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:913417 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	R2	chrUextra:9355577 (-)	GGGGAUC	113	GUCGACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGCGGC	
D. melanogaster	**R2**	**chrXHet:198910 (+)**	**GGGGAUC**	**113**	**GUCGACUCAAA**	**ACC**	**UCCUCGU**	**GGU**	**GGUCCCC**	**GGUAA**	**18**	**GCUAA**	**UUUGAGCGGC**	drz-Dmel-1-2
D. sechellia	R2	super_1387 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. sechellia	R2	super_1786:3915 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. sechellia	R2	super_2180:400 (+)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. sechellia	R2	super_294:10603 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. sechellia	R2	super_4663:1872 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. sechellia	**R2**	**super_700:5309 (-)**	**GGGGAUC**	**112**	**GUCGACUCAAA**	**ACC**	**UCCUCGU**	**GGU**	**GGUCCCC**	**GGUAA**	**18**	**GCUAA**	**UUUGAGCGGC**	drz-Dsec-1
D. sechellia	R2	super_757:5113 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. sechellia	R2	super_8043:867 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. sechellia	R2	super_8352:1206 (+)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. sechellia	R2	super_971:275 (+)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:12442859 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:15080282 (+)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:15188210 (+)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:1918327 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:3651895 (+)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:6203424 (+)	GGGGAGU	89	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:6302210 (+)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:9223417 (-)	GGGGAUC	116	ACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUAA	18	GCUAA	UUUGAGC	
D. simulans	R2	chrU:9334398 (+)	GGGGAAU	39	GACUCAAA	ACC	UCCUCGU	GGU	GGUCCCC	GGUA	76	UCAUA	UUGGAAC	
D. simulans	**R2**	**chrX_random:6649216 (-)**	**GGGGAGU**	**85**	**GUCGACUCAAA**	**ACC**	**UCCUCGU**	**GGU**	**GGUCCCC**	**GGUAA**	**18**	**GCUAA**	**UUUGAGCGGC**	drz-Dsim-1
			GAGGGAUC	**112**	**GUCGACUCAAA**	**ACC**	**UCCUCGU**	**GGU**	**GGUCCCC**	**GGUAA**	**18**	**GCUAA**	**UUUGAGCGGC**	
D. yakuba	R2	chrU:15871674 (+)	GGGGAAC	74	ACUCAAA	ACC	UCCUCGU	GGU	GUUUCCC	GGUAA	18	GCUAA	UUUGAGC	
D. yakuba	R2	chrU:18155939 (-)	GGGGAAC	73	ACUCAAA	ACC	UCCUCGU	GGU	GUUUCCC	GGUAA	18	GCUAA	UUUGAGC	
D. yakuba	R2	chrU:24016810 (+)	GGGGAAC	73	ACUCAAA	ACC	UCCUCGU	GGU	GUUUCCC	GGUAA	18	GCUAA	UUUGAGC	
D. yakuba	R2	chrU:25452614 (+)	GGGGAAC	73	ACUCAAA	ACC	UCCUCGU	GGU	GUUUCCC	GGUAA	18	GCUAA	UUUGAGC	
D. ananassae	**R2**	**scaffold_13163:86749 (+)**	**GGAGAAU**	**62**	**GUCGACUCAAA**	**ACC**	**UCCUCGU**	**GGU**	**AUUCUCU**	**GGG**	**18**	**GCUGA**	**UUUGAGCGGC**	drz-Dana-1

Species	Element	Location	Seq1	N1	Seq2	Seq3	Seq4	Seq5	Seq6	Seq7	N2	Seq8	Seq9	Name
D. yakuba	R2B	chrU:23410395 (-)	GGGGAAC	80	ACUCAAA	ACC	UCCUCGU	GGU	GUUUCCC	GGUAA	18	GCUAA	UUUGAGC	
D. persimilis	**R2B**	**super_2450:4278 (-)**	**GGAAGAU**	**70**	**GAUUCAAG**	**ACC**	**UCCUCGU**	**GGU**	**AUCUUCU**	**GGA**	**35**	**ACUAA**	**CUUGAAC**	**drz-Dper-3**
D. persimilis	**Baggins**	**super_43:160102 (+)**	**GGCCGUC**	**96**	**CGACUAUA**	**CUC**	**UCCCCGU**	**GAG**	**GGCGGCU**	**GGAAA**	**37**	**GCGAA**	**UGGUAGUCG**	**drz-Dper-2**
D. ananassae	Baggins	scaffold_622:299 (+)	GGCCGUC	70	CGACUAUA	CUC	UCCACGU	GAG	GGCGGCU	GGUAA	38	GCGAAU	CGUAGUCG	
D. melanogaster	Baggins	chr2L:22970826 (+)	GGCCGCC	73	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	43	GCUAAU	CGUAGCUG	
D. melanogaster	Baggins	chr2R:697370 (-)	GGACACC	80	AACUUCA	CUC	UCCACGC	GAG	GGUGGCU	GGA	169	GCUAAC	GGACAGUC	
D. melanogaster	Baggins	chr2R:721076 (+)	GGCCGCA	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	12	GCUAAU	CGUAGUUG	
D. melanogaster	Baggins	chr2RHet:1908889 (-)	GCUUCA	250	GGCGGAU	CUC	UCCCCGAC	GAG	GGAAGU	GGG	152	CCAUAA	AGGCGCU	
D. melanogaster	Baggins	chr2RHet:2197264 (-)	GGCCGCC	69	ACUGCA	CUC	UCCACGC	GAG	GGCGGCU	GAU	193	GCUAAU	GGAGAGU	
D. melanogaster	**Baggins**	**chr3LHet:191774 (+)**	**GGCCGCC**	**79**	**CAACUACA**	**CUC**	**UCCACGC**	**GAG**	**GGCGGCU**	**GGAAA**	**34**	**GCUAAU**	**CGUAGUUG**	**drz-Dmel-2-3**
D. melanogaster	Baggins	chr3LHet:2034646 (-)	GGCCGCU	81	ACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGA	169	GCUAAC	GGAGAUU	
D. melanogaster	Baggins	chr3LHet:832291 (+)	GGCCACC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGA	96	CCGGAC	UAAGGUG	
D. melanogaster	**Baggins**	**chrU:1256859 (-)**	**GGCCGCC**	**50**	**CAACUACA**	**CUC**	**UCCACGC**	**GAG**	**GGCGGCU**	**GGGAA**	**39**	**GCUAAU**	**CAUUGUUG**	**drz-Dmel-2-1**
D. melanogaster	**Baggins**	**chrU:2434561 (-)**	**GGCCGCC**	**79**	**CAACUACA**	**CUC**	**UCCACGC**	**GAG**	**GGCGGCU**	**GGAAA**	**39**	**GCUAAU**	**CGUAGUUG**	**drz-Dmel-2-2**
D. melanogaster	Baggins	chrU:6970241 (+)	GGCCGCC	74	CAACUACA	CUC	UCCGCGC	GAG	GGCGGCU	GGAA	27	GCUAAG	AGUGCUG	
D. melanogaster	Baggins	chrXHet:161365 (-)	GACCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. sechellia	Baggins	super_6206:1441 (-)	GGCCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAAC	38	GCUAAU	CGUUGUUG	
D. sechellia	Baggins	super_63:22187 (-)	GGCUGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. sechellia	Baggins	super_72:87870 (-)	GGCCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. sechellia	Baggins	super_731:795 (-)	GGCCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. sechellia	Baggins	super_731:5452 (-)	GGCCGCC	79	AAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. sechellia	Baggins	super_777:799 (+)	GGCUGCC	79	CAAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	CCUAAU	CGUAGUUG	
D. sechellia	Baggins	super_921:11825 (-)	GGUCGCC	78	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. sechellia	Baggins	super_99:64843 (-)	GGGCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. simulans	Baggins	chr2h_random:992223 (+)	GGCCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCCAAU	CGUAGUUG	
D. simulans	Baggins	chr2L:16805900 (-)	GGCCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	37	GCUAAU	CGUAGUUG	
D. simulans	Baggins	chr2R_random:548618 (+)	GGCCGCC	79	CAACUGCA	CUC	UCCACGC	GAG	GGCGGCU	GGGAA	39	GCUAAU	CGUAGUUG	
D. simulans	Baggins	chr2R_random:562083 (-)	GGCCGCC	79	CAACUGCA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. simulans	Baggins	chr3h_random:1377615 (-)	GGCCGCC	66	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. simulans	Baggins	chrU:14353216 (+)	GGCCGCC	73	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	
D. simulans	Baggins	chrU:1800152 (-)	GGCCGCC	79	CAACUACA	CUC	UCCACGC	GAG	GGCGGCU	GGAAA	39	GCUAAU	CGUAGUUG	